

Análisis de variables ordinales en modelos de variable latente:  
evaluación de métodos de estimación y procedimientos de ajuste

Edixon Chacón

Universidad Autónoma de Madrid  
Departamento de Psicología Social y Metodología

Análisis de variables ordinales en modelos de variable latente:  
evaluación de métodos de estimación y procedimientos de ajuste

Tesis doctoral:

Autor: Edixon Chacón

Directores:

Dr. Jesús M. Alvarado I.

Dra. Carmen Santisteban R.

2008

# ÍNDICE

<b>INTRODUCCIÓN .....</b>	<b>1</b>
---------------------------	----------

## **CAPITULO I.**

### **Modelos de variables latentes, características y clasificación**

1.1. Variables latentes y modelos de variables latentes .....	11
1.2. Clasificación de los LVM .....	11
1.3. Modelo general del análisis factorial .....	13
1.4. Análisis factorial para datos ordinales .....	14
1.4.1. Aproximación de underlying variables (UV).....	15
1.4.2. El análisis factorial y la teoría de respuesta al ítem .....	16
1.5. Modelo general de variables latentes .....	19
1.5.1. Estimación y ajuste de LVM.....	21
1.5.2. Máxima verosimilitud conjunta (JML) .....	23
1.5.3. Máxima verosimilitud marginal (MML).....	23
1.5.4. Máxima verosimilitud condicional (CML) .....	24
1.5.5. Máxima verosimilitud información completa (FIML).....	25
1.6. Modelos de rasgo latente (LTM): aproximación UV .....	26
1.7. Modelos de rasgo latente (LTM): aproximación IRT.....	28
1.7.1. Respuestas binarias .....	28
1.7.2. Respuestas ordinales .....	29
1.8. Conclusiones del capítulo .....	29

## **CAPITULO II.**

### **Tratamiento de datos ordinales en modelos de rasgo latente**

#### **(LTM)**

2.1. Introducción .....	32
2.2. Aproximaciones teóricas.....	33
2.2.1. Aproximación UV .....	33
2.2.2. Aproximación IRT .....	34
2.2.3. Estandarización de parámetros.....	38
2.2.4. Ajuste del modelo .....	39
2.3. Comparación de métodos de estimación.....	40
2.3.1 Problemas resueltos y no resueltos .....	40
2.3.2 Una solución aproximada.....	40
2.4. Método .....	41
2.4.1. Aproximación empírica.....	41
2.4.1.1. Sujetos .....	41
2.4.1.2. Instrumento.....	41
2.4.1.3. Software .....	43
2.4.1.4. Estimación y ajuste de datos empíricos .....	43
2.4.2. Simulación Monte Carlo .....	51
2.4.2.1 Resultados de la simulación .....	52
2.4 Discusión y conclusiones .....	55

## **CAPITULO III.**

### **Evaluación del ajuste y procedimientos de simulación en LTM**

3.1. Introducción .....	56
3.2. Una solución aproximada .....	58
3.3. Método .....	60
3.3.1. Bootstrap .....	60
3.3.2. Monte Carlo .....	63

3.3.3. Simulación Monte Carlo para datos ordinales .....	64
3.4. Estudio de simulación .....	63
3.5. Resultados .....	68
3.6. Discusión y conclusiones .....	74

## **CAPITULO IV. Aportaciones en aspectos técnicos y computacionales para la evaluación del ajuste en LTM**

4.1. Introducción .....	77
4.2. Procedimiento .....	79
4.3. Implementación.....	80
4.4. Aplicación .....	83
4.4.1. Muestra.....	83
4.5. Resultados .....	83
4.6. Conclusiones .....	87
Apéndice .....	88

## **CAPITULO V: Un nuevo test para evaluar la bondad de ajuste en tablas de contingencia de amplia dispersión. Comparación con otros índices.**

5.1. Introducción .....	90
5.2. Problemas resueltos y no resueltos .....	91
5.3. Una solución aproximada .....	93
5.4. Método .....	93
5.4.1. Ajuste en tablas de contingencia de amplia dispersión.....	93
5.4.2. Reglas de interpretación.....	95
5.4.3. Sujetos .....	95
5.4.4. Instrumento .....	95

5.4.5. Software .....	96
5.5. Simulación .....	96
5.6. Resultados de la simulación.....	97
5.7. Formulación de los nuevos coeficientes .....	99
5.8. Aplicación .....	100
5.9. Conclusiones .....	106
Apéndice .....	108
 CONCLUSIONES .....	 111
REFERENCIAS BIBLIOGRÁFICAS.....	117
ANEXOS .....	129
Tablas .....	130
Figuras.....	132
Acrónimos .....	133
Notación .....	134

# **Introducción**

El tratamiento de datos ordinales es una situación con la que se enfrentan constantemente los investigadores en las ciencias sociales. Tradicionalmente su análisis se abordaba en muchas ocasiones con las mismas técnicas usadas para los datos continuos, aunque siempre se han propuesto métodos que han sido útiles en casos particulares. En la actualidad, en el campo de la medición psicológica, es de especial interés la aplicación de modelos de variables latentes (LVM) para abordar los análisis con datos ordinales. Bajo este enfoque de las LVM, el investigador dispone de una variedad de métodos, técnicas y programas informáticos entre los cuales debe elegir el que sea adecuado para aproximar los datos a los objetivos de su investigación.

En este conjunto de opciones hay aproximaciones teóricas y metodológicas especialmente relevantes. Habitualmente, destacan los enfoques clásicos que consisten en tratar a los datos ordinales como variables subyacentes (underlying variables, UV), en los que se establece una conexión entre variables categóricas y continuas. Una alternativa emergente consiste en aproximar los LVM desde la teoría de respuesta al ítem (IRT) que proporciona un conjunto

adicional de modelos, métodos de estimación y de evaluación del ajuste, que implican la asunción de nuevos retos y supuestos.

Entre los temas más destacados y que plantean una mayor problemática en los modelos es la estimación de sus parámetros y su validez. Por ello, hay gran interés y se está desarrollando mucha investigación en las fases de estimación y la evaluación del ajuste en LVM. Los investigadores han conducido sus hipótesis a indagar acerca de las posibilidades y potencialidades de métodos y técnicas que sean congruentes con la naturaleza de los datos. El origen de las propuestas está cada vez más diversificado atendiendo a las condiciones particulares de las investigaciones. Así pues, nos encontramos con métodos y modelos específicos según la naturaleza métrica o no de las variables latentes y de las observadas.

En muchas áreas aplicadas y en general en las ciencias sociales y del comportamiento conviene el estudio de las variables ordinales observadas, independientemente de la naturaleza (ordinal o no) de las variables latentes. Esto condiciona las opciones de análisis y establece una ramificación en esta área de estudio, al considerar dos ramas, los modelos de rasgo latente (LTM) y los modelos de clases latentes (LCM).

En el tratamiento del ajuste de LVM, y en especial en los LTM, se suscitan varias situaciones que despiertan un especial interés por la investigación, la evaluación de las diferentes alternativas tanto teóricas como metodológicas, la evaluación de los test de bondad de ajuste, así como los instrumentos que requieren la generación de nuevo software. Por lo tanto en esta tesis se plantean los temas que a continuación enumeramos en los epígrafes del 1 al 4,

## **1.- Evaluación de alternativas teóricas y metodológicas**

Habitualmente, en las investigaciones en las que se emplean modelos de variables latentes, como el análisis factorial (FA) y los modelos de ecuaciones estructurales (SEM), los datos ordinales se tratan como datos continuos. Esto



implica que se asumen los mismos principios y supuestos que se aplican para datos continuos. Uno de los supuestos que merece especial atención es el de la normalidad, pues por ejemplo, en los cuestionarios que miden actitudes generalmente las respuestas se dan en una escala tipo Likert, y asumir que esos datos se pueden distribuir como una distribución normal multivariante es un requisito difícil de cumplir. Esta situación condiciona las expectativas del investigador con respecto a la formulación y prueba de sus modelos.

Asimismo, en muchos trabajos no se aplican los nuevos desarrollos para el tratamiento de datos ordinales y su relación con los modelos de variables latentes, lo cual no permite aprovechar las posibilidades de distintos enfoques teóricos que conllevan a resultados más rigurosos, como es el caso del tratamiento de estos modelos desde la IRT. Esto trae como consecuencia que no se precise en la eficiencia entre distintos métodos de estimación y evaluación del ajuste.

Por otra parte, los estudios con variables ordinales se centran en la comparación de los parámetros, obtenidos por diferentes métodos de estimación, en las que básicamente se estudia la precisión de las estimaciones y apenas se abordan los efectos que tiene el método de estimación empleado en el ajuste de los modelos.

Expuestos los anteriores planteamientos cabe ahora formular algunas interrogantes: al ajustar modelos con datos ordinales ¿Se obtienen resultados similares empleando métodos de estimación desde distintas aproximaciones teóricas? ¿Cómo afecta el método de estimación adoptado a la bondad del ajuste del modelo? ¿Qué alternativas metodológicas existen actualmente para un tratamiento más eficaz en el análisis de datos ordinales? ¿Ofrecen mejores resultados los métodos lineales o los basados en la IRT?

## **2.- Evaluación de los factores que afectan el ajuste de los modelos**

Existen varios factores que afectan la bondad del ajuste del modelo. Diversos estudios de simulación han evaluado el comportamiento de los índices de ajuste y han determinado cómo estos índices son afectados por características tales como el número de factores, el número de indicadores por factor y el tamaño de las cargas factoriales (Anderson & Gerbing, 1984; Bearden, Sharma & Teel, 1982; Boomsma, 1982).

Las causas que afectan el ajuste, en LVM, han sido abordadas principalmente desde los SEM, pero escasamente han sido evaluados desde la IRT. El procedimiento de ajuste, entendido como las etapas de estimación de parámetros y evaluación de los índices del modelo, debe ser tratado con madurez y rigor. Un enfoque especial para validar este procedimiento lo constituyen los procedimientos de simulación y remuestreo. En este contexto un nuevo obstáculo se presenta a los investigadores, la elección de un procedimiento de simulación para generar muestras y la utilización de un enfoque paramétrico o no de las mismas.

Con propósitos similares, pero con distinta fundamentación e implementación, los procedimientos de simulación y remuestreo presentan diversos escenarios y posibilidades en el ajuste de datos ordinales. En la mayoría de los casos, los investigadores confían en el software comercial cuando utilizan técnicas o adoptan métodos que permiten evaluar el ajuste sin indagar en los detalles del programa que da origen a los resultados que posteriormente interpretan. Las técnicas de simulación, como Monte Carlo y Bootstrap, han sido empleadas ampliamente para verificar la calidad de las estimaciones y validar el ajuste en los LVM, pero no se ha evaluado el efecto que tiene la elección de una u otra técnica de simulación sobre el ajuste del modelo.

En el primer epígrafe nos preguntamos ¿Cómo afecta el método de estimación a la bondad de ajuste del modelo? A esta pregunta ahora podemos añadir ¿Se obtienen resultados similares en el ajuste del modelo, al adoptar Monte Carlo o Bootstrap, para validar los índices de ajuste? ¿Cuál es el efecto conjunto

que tienen el método de estimación adoptado y la técnica de simulación seleccionada en el ajuste del modelo?

### **3.- Consideración de los aspectos técnicos y computacionales: posibilidades de las herramientas de software empleadas**

El uso de herramientas de software constituye un elemento importante en el análisis y obtención de los resultados. En muchos casos estas herramientas no satisfacen las necesidades del investigador, por lo que se hace necesario buscar caminos alternativos.

Las nuevas aproximaciones de los LVM basadas en IRT ofrecen escasas herramientas de software para el tratamiento de datos ordinales. Muchas de ellas ofrecen soluciones parciales y en algunos casos no están consolidadas, presentando errores al ejecutarlas. Algunos programas están diseñados para evaluar escalas unidimensionales, por lo que la evaluación de modelos multidimensionales de la IRT requiere la creación de programas propios o el empleo simultáneo de varias herramientas.

En el tratamiento de datos ordinales se atizan distintos procedimientos implementados en software comercial como AMOS, MPLUS, LATENTGOLD y LISREL, entre otros. En esta tesis hemos utilizado LISREL/PRELIS que tiene implementados programas que permiten el ajuste de modelos bajo la aproximación de las UV, así como comandos que permiten aproximar LVM desde la IRT usando métodos de estimación de información completa. Sin embargo, cuando se utilizan los procedimientos de simulación Bootstrap y Monte Carlo, este software presenta algunas limitaciones técnicas que impiden llevar a cabo la evaluación general de los modelos.

Para la evaluación del ajuste, en muchas ocasiones se combinan la estimación de los parámetros y el empleo de técnicas de simulación, por lo que

convendría disponer de una herramienta funcional que permita evaluar modelos unidimensionales y multidimensionales de la IRT, estimar sus parámetros con métodos de estimación de información completa, obtener sus índices de ajuste y validarlos a través de procedimientos de simulación ¿Qué ventaja y limitaciones presenta una herramienta con estas características? ¿Cuáles son los detalles de su implementación?

#### **4.- Selección de estadísticos de bondad de ajuste que permitan evaluar adecuadamente el ajuste de los modelos**

El tema del ajuste de modelos de variables no siempre resulta sencillo, y generalmente no se resuelve usando el estadístico chi-cuadrado. Los índices de bondad de ajuste absoluta como el  $\chi^2$  o  $G^2$ , no permiten distinguir causas de desajuste como podría ser la presencia de un ítem en particular que esté afectando el ajuste o que sea necesario añadir nuevos parámetros al modelo.

Se han propuesto varias técnicas y test de ajuste para evaluar LVM, basadas en el análisis de las distribuciones marginales, por pares o tríos de ítems. La reglas de interpretación del ajuste al usar estas técnicas no están sistematizadas, son imprecisas, resultan arbitrarias y poco funcionales. Se basan en la comparación del valor de los índices con un umbral establecido para considerar el ajuste como adecuado. Esta situación desconcierta a los investigadores a la hora de tomar decisiones acerca de la calidad del ajuste, pues la existencia de algún valor que supere este umbral es suficiente para cuestionar la solución y considerar el ajuste como inadecuado.

Por otra parte, se presenta la situación en tablas de contingencia de amplia dispersión (sparseness) con  $m^p$  patrones de respuesta en las que los índices existentes se pueden aplicar para evaluar la bondad del ajuste en cada una de las celdas, o para evaluar las distribuciones marginales de las tablas de contingencia para pares de variables, pero no dan información acerca de la evaluación global

del modelo. Esto es un gran inconveniente cuando hay que evaluar un número elevado de soluciones.

Con respecto a los test de bondad del ajuste en LTM surgen varias interrogantes ¿Cuál es el procedimiento más adecuado para evaluar el ajuste en tablas de contingencia de amplia dispersión? ¿Cómo se interpreta los valores de ajuste por pares de ítems y cómo afecta cada uno la evaluación global del modelo? ¿Se puede evaluar a partir de un valor único la bondad del ajuste del modelo? ¿Qué beneficios proporciona un índice que permita evaluar el ajuste global del modelo?

Para dar respuestas a estos planteamientos e interrogantes, en el presente trabajo se tratan diversos tópicos sobre datos ordinales en LVM, que incluyen las características y clasificación de los LVM, haciendo un especial énfasis en los datos ordinales y su relación con los LTM; aproximaciones teóricas y metodológicas para la estimación y el ajuste de datos ordinales en LTM; el efecto del procedimiento de simulación y los métodos de estimación en el ajuste de LTM; revisión de los aspectos técnicos y computacionales de las condiciones de simulación para evaluar el ajuste en LTM y una evaluación de los estadísticos para evaluar el ajuste en LTM.

Los objetivos propuestos son los siguientes:

1. Mostrar cómo el ajuste con ordinal data se ve afectado por el método de estimación elegido (GLS, WLS, NOR y POM) y determinar cuál es la mejor opción para ajustar estos modelos LTM
2. Explicar el efecto que pueden tener el procedimiento de simulación escogido para generar muestras (Bootstrap o Monte Carlo) y el método de estimación de información completa adoptado (POM o NOR) en los índices de ajuste en LVM aproximados desde la IRT.
3. Proponer un procedimiento de software que usando la simulación Monte Carlo permita el ajuste de modelos de variables latentes (LVM) bajo la aproximación IRT.

4. Proponer coeficientes globales, calculados a partir de la información suministrada por los índices de ajuste para tablas de contingencia de amplia dispersión con  $m^p$  patrones de respuesta, que permitan evaluar el ajuste en modelos de la IRT con métodos de estimación de información completa.

La tesis está estructurada en cinco capítulos. En el capítulo I se hace una breve introducción teórica acerca de las características y clasificación de los LVM y se examinan las distintas aproximaciones teóricas y metodológicas para el tratamiento de datos ordinales.

En el capítulo II se hace un estudio comparativo de la evaluación del ajuste usando las aproximaciones UV e IRT. Concretamente se evalúa el ajuste obtenido al adoptar métodos de estimación de información limitada como el GLS y el WLS y métodos de estimación de información completa (FIML), adoptando como funciones de enlace la logística (POM) y normal (NOR) en la IRT. Los resultados del ajuste son validados a partir de un estudio de simulación y se discute acerca de las implicaciones teóricas y metodológicas.

En el capítulo III se trata la medición del efecto que tiene la adopción de un procedimiento de simulación empleado para validar índices de ajuste cuando se usa en conjunto con un método de estimación específico (POM o NOR) en LTM tratados desde la IRT. Se presentan las causas producidas por estas variables y cómo estas afectan el ajuste y se discuten y contrastan los resultados con otras investigaciones.

En el capítulo IV se propone un procedimiento que permite generar muestras por simulación y evaluar los índices de ajuste en LTM desde la IRT. El capítulo describe las limitaciones y fallas que presenta el software LISREL/PRELIS y compara el procedimiento seguido por los autores Jöreskog & Sörbom (1996) para la aproximación UV con el procedimiento propuesto bajo la aproximación IRT.

En el capítulo V se presentan dos nuevos coeficientes de ajuste que permiten, a partir de un único valor, evaluar el ajuste global del modelo. Los coeficientes son obtenidos en el marco de la polémica del ajuste en tablas de contingencia de amplia dispersión (*sparseness*) para ítems politómicos. Se hace una revisión de los índices de ajuste que se aplican comúnmente en esta situación y aprovechando la información que suministran éstos índices, se formulan los nuevos coeficientes. Se presenta la relación teórica entre las medidas de ajuste existentes y los coeficientes planteados a través de una pirámide de ajuste en la que se resaltan las contribuciones y aplicaciones de las nuevas medidas y se muestra la aplicación de los nuevos coeficientes con datos empleados por otros autores y se discute acerca de sus implicaciones teóricas y metodológicas.

La resolución de los problemas planteados muestra una coherencia teórica y metodológica que permite tratar un tema común: la estimación y ajuste en modelos de rasgo latente. Aunque en esta investigación sólo se está abordando una pequeña porción del amplio espectro en esta materia, se considera que el tratamiento de enfoques tradicionales y emergentes en tema del ajuste de estos modelos es abordado con seriedad y que las nuevas contribuciones expuestas repercuten positivamente al enriquecimiento teórico y metodológico de esta área del conocimiento.

# Capítulo I

## Modelos de variables latentes, características y clasificación

### Resumen

Este capítulo introduce diversos tópicos referentes a la naturaleza y características de los modelos de variables latentes (LVM), se hace un especial énfasis en los modelos de rasgo latente (LTM) basados en datos ordinales. La explicación intuitiva del concepto de variable latente abre paso a su clasificación y a la justificación de su uso en diversas áreas. Se explora en los inicios del análisis factorial introducido por Spearman (1904), y a partir de este, se explica brevemente la obtención de otros modelos de variables latentes desde la aproximación lineal del análisis factorial y desde la teoría de respuesta al ítem (IRT). Se describe brevemente el modelo general de variables latentes y se exploran los diversos métodos de estimación de parámetros y en el ajuste de estos modelos. Finalmente se exponen los LTM para datos binarios y ordinales.



### **1.1 Variables latentes y modelos de variables latentes (LVM)**

Ciertos conceptos en las ciencias sociales y del comportamiento no son fáciles de definir y existe mucha discusión acerca de su real significado, por ejemplo: clase social, opinión pública o personalidad extrovertida. A estos conceptos comúnmente se les llama variables latentes, pues no se pueden observar directamente en una población. Se consideran esencialmente constructos hipotéticos inventados por los científicos con el propósito de comprender un área de investigación de interés y para los cuales no existe métodos operativos para medirlos directamente (Everitt, 1984).

Aunque las variables latentes no son observadas el efecto, que ejercen sobre ella la medida de otras variables si observadas (o manifiestas) sí lo es. Se trata pues de explicar a partir del estudio de las relaciones entre variables observadas un número menor de variables latentes, a partir de métodos o modelos específicos que permitan evaluar la adecuación de las primeras.

Un modelo estadístico especifica la distribución conjunta de un conjunto de variables aleatorias y si se trata de un LVM algunas de esas variables (las variables latentes) no son observadas. Un LVM es por lo tanto una expresión matemática que resume estas relaciones entre variables latentes y observadas. El estudio de variables latentes es un tema de interés en múltiples áreas: psicología, sociología, economía, entre otras. Bartholomew et al., 2002 describe ampliamente su relación con la medición de constructos ampliamente estudiados como la inteligencia, las actitudes políticas o el status socioeconómico.

Hasta hace un par de décadas, el método clásico por excelencia para analizar variables latentes fue el análisis factorial (FA), en Bartholomew et al., 2002, se plantea una evolución de los modelos de variables latentes, en donde se distingue entre las viejas aproximaciones: FA, modelos de clases latentes (LCM), modelos de rasgo latente (LTM) y las nuevas aproximaciones basadas en enfoque bayesianos.

### **1.2 Clasificación de los LVM**

La naturaleza de la escala en la que se miden las variables manifiestas y latentes (nominal, ordinal, intervalo o de razón), da origen a una clasificación de los modelos de variables latentes. Bartholomew & Knott (1999), clasifican los LVM en función de la escala de las variables latentes y observadas, distinguiendo entre cuatro tipos (Tabla 1.1). Las denominadas variables métricas toman valores en un conjunto de números y pueden ser discretas o continuas. Las variables categóricas asignan una categoría entre un conjunto de categorías, que pueden estar ordenadas o no. Si se asume orden en las categorías pueden ser tratadas como variables métricas.

El objetivo principal de estos modelos es reducir la complejidad de un conjunto de datos explicando la asociación entre variables observadas en términos de un conjunto más pequeño de factores (FA), clases latentes (LCM), rasgos (LTM) o perfiles (LPM).

Tabla 1.1

Clasificación de los LVM según la métrica de variables latentes y observadas

		Variables manifiestas	
		Métricas	Categóricas
Variables	Métricas	Análisis Factorial (FA)	Modelos de rasgos latentes (LTM)
	Categóricas	Modelos de perfiles latentes (LPM)	Modelos de clases latentes (LCM)

En el AF y LTM, las variables latentes son tratadas como variables continuas normalmente distribuidas. Por otra parte, en LPA y LCM, las variables latentes son discretas y por lo tanto se asume que provienen de una distribución binomial o multinomial. Las variables manifiestas en AF y LPA son continuas, en este caso se asume que sus distribuciones condicionales son normales. En LTM y

LCM, los indicadores o variables manifiestas son categóricos (binarios o ordinales) y se asume que sus distribuciones condicionales son binomiales o multinomiales. Algunas investigaciones han demostrado que la distribución de un modelo de variables latente continuas puede ser aproximado por una distribución discreta (Heinen, 1996), esto evidencia que, en algunos casos, la distinción entre variables latentes discretas y continuas puede carecer de importancia.

Las funciones que muestran la distribución de los indicadores en LVM están relacionadas con la naturaleza de las escalas en que son medidos, si la variable latente es tratada como discreta o continua se pueden obtener formas generalizadas para los LTM y LCM, respectivamente, diversificando así su desarrollo y aplicaciones (Bartholomew et al., 2002)

### 1.3 Modelo general del análisis factorial

El análisis factorial fue introducido por Spearman (1904). La formulación general de ésta técnica, considera un conjunto  $p$  de variables observadas o manifiestas  $x$ s que se resumen en  $q$  factores o variables latentes no observadas  $y$ s, ( $q < p$ ), siguiendo el modelo de *regresión múltiple*, la expresión que relaciona cada variable observada como una combinación lineal de los factores, viene dada por:

$$x_i = \alpha_{i0} + \alpha_{i1}y_1 + \alpha_{i2}y_2 + \dots + \alpha_{iq}y_q + e_i \quad (i = 1 \dots p) \quad (1)$$

Donde  $y_1, y_2, \dots, y_q$  son los factores comunes o variables latentes,  $e_1, e_2, \dots, e_p$  son los residuales o términos de error,  $\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iq}$  son las cargas factoriales y  $\alpha_{i0}$  representa el término constante o intercepto, la ecuación 1 representa el *modelo general del análisis factorial lineal*.

Aunque el análisis factorial lineal se basa en la idea de la regresión múltiple, en lugar de tener una variable dependiente se tienen  $p$  variables

(observadas), esto hace más compleja la comprobación de supuestos. Se debe de verificar que las variables manifiestas no estén correlacionadas, y que además, estén explicadas a partir de las variables no observadas o factores. Bartholomew et al., 2002, resume estos supuestos así:

- i. Las variables latentes o factores  $y_1, y_2, \dots, y_q$  no están correlacionados.
- ii. Cada variable latente  $y_1, y_2, \dots, y_q$  se distribuyen normalmente con media cero y varianza uno.
- iii. Los términos de error  $e_1, e_2, \dots, e_p$  no están correlacionados.
- iv. Cada término de error se distribuyen normalmente con media cero pero puede tener diferente varianza  $Var(e_i) = \sigma_i^2, (i = 1 \dots p)$ .
- v. Las variables latentes  $y_1, y_2, \dots, y_q$  y los errores  $e_1, e_2, \dots, e_p$  no están correlacionados.
- vi.  $y_1, y_2, \dots, y_q$  siguen una distribución normal multivariante.
- vii.  $e_1, e_2, \dots, e_p$  siguen una distribución normal multivariante.

Los supuestos *vi* y *vii* implican que  $x_1, x_2, \dots, x_p$  siguen una distribución normal multivariante y esto conduce a un *modelo lineal normal del análisis factorial*. Los supuestos *iii* y *v* implican que las correlaciones entre las variables observadas  $x$ s son explicadas suficientemente por los factores.

#### 1.4 Análisis factorial para datos ordinales

El modelo lineal general resulta inadecuado cuando las variables observadas son ordinales, debido a que en este modelo las variables manifiestas se asumen como continuas. El problema se reduce a una violación de los supuestos básicos, puesto que si tenemos que los errores se distribuyen normalmente con media 0 y varianza  $\sigma_i^2$  (supuesto *iv*) y las variables siguen una distribución normal estándar  $y_j \sim N(0,1)$  para todo  $j$  (supuesto *ii*), el vector de variables latentes  $\mathbf{y}=(y_1, y_2, \dots, y_q)$  y los términos de error  $e_i$  pueden tomar valores

independientes, y por lo tanto  $x_i$ , también puede tomar cualquier valor. Esto es lo que invalida al modelo factorial lineal para variables categóricas (binarias y ordinales). En consecuencia, es necesario un modelo diferente que permita relacionar el vector de variables latentes  $y$  con las variables observadas. Se han adoptado principalmente dos aproximaciones teóricas para tratar este problema.

#### 1.4.1 Aproximación de *underlying variables* (UV)

Una primera aproximación es el enfoque de *underlying variables* (UV), en el cual a partir de las variables categóricas observadas  $x_i$  se generan variables continuas ficticias  $x_i^*$ , consideradas UV, éstas variables no son observadas y no deben confundirse con las variables latentes, se les puede denominar *variables observadas incompletas* pues su valor depende, de si exceden o no, los valores de los umbrales (Bartholomew et al., 2002).

Se asume que, para cada variable observada  $x_i$ , existe una variable observada incompleta continua  $x_i^*$ , la cual se distribuye normalmente con media  $\mu_i$  y varianza  $\sigma_i^2$ . La conexión entre  $x_i$  y  $x_i^*$  está basada en los umbrales  $\tau_i$ , si  $x_i^*$  se encuentra por debajo de un determinado umbral  $\tau_i$ ,  $x_i$  toma un valor en las categorías (0,1,2,3...). La esencia de este método es tratar las  $x_i^*$  bajo el modelo lineal general del análisis factorial, esto supone que:

$$x_i^* = \alpha_{i0}^* + \alpha_{i1}^* y_1 + \alpha_{i2}^* y_2 + \dots + \alpha_{iq}^* y_q + e_i \quad (i = 1 \dots p) \quad (2)$$

Donde  $y_1, y_2, \dots, y_q$  son los factores comunes o variables latentes,  $e_1, e_2, \dots, e_p$  son los residuales o términos de error,  $\alpha_{i1}^*, \alpha_{i2}^*, \dots, \alpha_{iq}^*$  son las cargas factoriales y  $\alpha_{i0}^*$  representa el término constante o intercepto. En el modelo

general del análisis factorial lineal, las  $x_i^*$  son variables observadas, en esta aproximación se consideran UV.

El ajuste del modelo lineal general se hace a partir de la matriz de correlaciones, por lo que es necesario calcular la matriz de correlaciones tetracóricas (datos binarios) o policóricas (datos ordinales). Para la obtención de estas matrices, previamente se calculan los umbrales a partir de las distribuciones marginales univariadas de las  $x_i^*$ , y las correlaciones se obtienen a partir de las distribuciones marginales bivariadas de las  $x_i^*$  con cada uno de los umbrales.

Los supuestos de la aproximación UV son resumidos en Bartholomew et al., 2002 como sigue:

- i. Las variables latente  $y_1, y_2, \dots, y_q$  son independientes y se distribuyen normalmente con media cero y varianza uno
- ii. Los errores son independientes y se distribuyen normalmente con media cero pero puede tener diferente varianza  $\sigma_i^2$ .
- iii. Las distribuciones univariadas y bivariadas de las  $x_i^*$  son normales

#### **1.4.2 El análisis factorial y la teoría de respuesta al ítem**

Una segunda aproximación se da desde la teoría de respuesta al ítem (TRI). En el caso de ítems binarios, el objetivo consiste en construir un modelo que exprese la probabilidad de una respuesta positiva  $\pi_i(y)$ , dada por un individuo seleccionado al azar, como una función de las variables latentes (Bartholomew et al., 2002). La probabilidad de una respuesta negativa es el complemento de las  $\pi_i(y)$ s. En el caso ordinal, donde existen más de dos categorías, es necesario especificar la probabilidad para cada categoría. Las variables observadas ordinales son denotadas por  $x_1, x_2, \dots, x_p$ . Suponemos que hay  $m_i$  categorías para la variable  $i$  que va desde  $1, \dots, m_i$ . En el caso binario  $m_i=2$ , las etiquetas pueden ser 0 y 1, pero pueden ser también 1 y 2. Para definir las

probabilidades de respuesta en cada categoría consideramos que  $\pi_{i(s)}(y)$  es la probabilidad de que dado  $y$ , la respuesta se encuentre en la categoría  $s$  para la variable  $i$ . Esta relación entre categoría y modelos se muestra a continuación:

Categorías	0	1				
Probabilidad de respuesta	$1 - \pi_i(y)$	$\pi_i(y)$				
Categorías	1	2	...	s	...	$m_i$
Probabilidad de respuesta	$\pi_{i(1)}(y)$	$\pi_{i(2)}(y)$	...	$\pi_{i(s)}(y)$	...	$\pi_{i(m_i)}(y)$

En ambos casos la suma de las probabilidades es 1. El caso binario, puede ser representado con el modelo logit, el cual está expresado como la probabilidad de obtener una respuesta de uno como una función lineal de las variables  $y$ :

$$\log it \pi_i(y) = \log_e \frac{\pi_i(y)}{1 - \pi_i(y)} = \alpha_{i0} + \sum_{j=1}^k \alpha_{ij} y_j \quad (3)$$

El problema se centra ahora en la generalización del modelo usado al caso de más de dos categorías. Partiendo de que el modelo logit (ecuación 3), puede ser aplicado para el caso de ítems politómicos, dividimos las categorías en dos grupos: las categorías  $(1, 2, \dots, s)$  en el primer grupo y las categorías  $(s+1, s+2, \dots, m_i)$  en el otro, el objetivo es conocer en cuál de los dos grupos se encuentra la respuesta. Las probabilidades de que una respuesta se encuentre en un grupo u otro, pueden escribirse como:

$$\gamma_{i(s)}(y) = \Pr(x_i \leq s) = \pi_{i(1)}(y) + \pi_{i(2)}(y) + \dots + \pi_{i(s)}(y)$$

y

$$1 - \gamma_{i(s)}(y) = \Pr(x_i > s) = \pi_{i(s+1)}(y) + \pi_{i(s+2)}(y) + \dots + \pi_{i(m_i)}(y)$$

Donde  $x_i$  denota la categoría dentro de la cual se encuentra la  $i$ -ésima variable. La probabilidades  $\gamma_{i(s)}(y)$  son conocidas como probabilidad de respuesta acumulada (cumulative response probabilities).

Podemos definir el modelo suponiendo que el modelo logit binario se sostiene para todas las divisiones de las posibles categorías  $m_i$  en dos grupos. Este modelo lo podemos formular de dos maneras equivalentes de acuerdo con el grupo que consideremos como “respuesta positiva”, es decir, en términos de  $\text{logit} \gamma_{i(s)}(y)$  o de  $\text{logit}(1 - \gamma_{i(s)}(y))$ . Lo común es hacerlo en función de  $(1 - \gamma)$  para relacionarlo directamente con el modelo binario donde  $(s=1, \dots, m_i; i=1, \dots, p)$ :

$$\log \left[ \frac{(1 - \gamma_{i(s)}(y))}{\gamma_{i(s)}(y)} \right] = \alpha_{i(s)} + \sum_{j=1}^q \alpha_{ij} y_j \quad (4)$$

En lugar de usar la función de enlace logit, se puede usar la función normal (probit) Si el modelo usa la función logit como función de enlace, también puede ser llamado proportional odds model (POM). El nombre de POM proviene del hecho, de que en el caso de un factor, la diferencia entre dos logit acumulativos para dos personas con puntuaciones factoriales  $y_1$  and  $y_2$  es proporcional a  $y_1 - y_2$ , por lo tanto, existe un parámetro intercepto  $\alpha_{i(s)}$  para cada categoría. Esto refleja el hecho de que al aumentar los umbrales ( $\alpha_{i(s)}$ ) para una respuesta dada la dificultad del ítem también aumenta. Un orden en las categorías implica que los parámetros de los interceptos también están ordenados:

$$\alpha_{i(1)} < \alpha_{i(2)} < \dots < \alpha_{i(m_i)}$$

Sin embargo, las cargas factoriales (que se corresponden con el parámetro  $\alpha$ .) permanecen iguales entre las categorías de la misma variable, en otras palabras el poder discriminatorio de los ítems no depende de la distancia, entre dos categorías adyacentes. Las  $\pi_s$  son obtenidas desde las  $\gamma_s$  por:



$$\pi_{i(s)}(y) = \gamma_{i(s)}(y) - \gamma_{i(s-1)}(y) \quad (s = 2, \dots, m_i)$$

Donde:  $\gamma_{i(1)}(y) = \pi_{i(1)}(y)$  y  $\gamma_{i(m_i)}(y) = 1$ . Por lo general, se hace referencia a  $\gamma_{i(s)}(y)$  como función de respuesta acumulada y a  $\pi_{i(s)}(y)$  como función de respuesta de categoría.

Los supuestos bajo la aproximación IRT, los cuales son comunes a los modelos factoriales, son los siguientes:

- i. Las variables latente  $y_1, y_2, \dots, y_q$  son independientes y se distribuyen normalmente con  $\mu = 0$  y  $\sigma^2 = 1$ .
- ii. Las respuestas de las variables observadas  $p$  tienen independencia condicional con las variables latentes  $y$

Con respecto al supuesto *i*, escoger la distribución normal para las variables latentes tiene ventajas durante la rotación de los factores, por fortuna la forma de la distribución de las variables latentes no tiene mucha influencia en el análisis e interpretación de los resultados (Bartholomew et al., 2002). Con el supuesto *ii* se pretende que las variables latentes  $y$  den cuenta de todas las asociaciones entre las variables observadas  $p$ , este supuesto de independencia condicional puede ser probado de forma indirecta verificando si el modelo ajusta a los datos, en un modelo de variables latentes, si el ajuste es bueno, las variables latentes explican la mayor parte de las asociaciones entre variables observadas, (Bartholomew et al., 2002).

### 1.5 Modelo general de variables latentes

Sea el vector columna  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$  un conjunto de  $p$  variables manifiestas y  $\mathbf{y}' = (y_1, y_2, \dots, y_q)$  las variables latentes, al igual que en el análisis factorial,  $q < p$ . Ambos conjuntos de variables, por definición, varían de un individuo a otro y teóricamente son representadas como variables aleatorias, por lo tanto, la relación entre ellas debe ser representada en términos de distribuciones

de probabilidad, así que después de que la información las variables  $x$  ha sido observada, la información que tenemos acerca de  $y$  es proporcionada por su distribución condicional dado  $x$  (Bartholomew & Knott, 1999).

En esencia todos los modelos de variables latentes asumen que  $x_1, x_2, \dots, x_p$  tienen una distribución de probabilidad condicional conjunta sobre  $y_1, y_2, \dots, y_q$  que denotaremos como:

$$\phi(\mathbf{x}|\mathbf{y})$$

si las variables manifiestas son continuas,  $\phi$  es una función de densidad, pero si es discreta, es un conjunto de probabilidades. Si la función de densidad de  $y$  es  $h(y)$  entonces la densidad incondicional de  $x$  esta dada por:

$$f(\mathbf{x}) = \int \phi(\mathbf{x}|\mathbf{y})h(\mathbf{y})d\mathbf{y} \quad (5)$$

El interés principal consiste en poder determinar  $y$  después de que  $x$  se ha observado. Esta información se obtiene por la densidad condicional:

$$h(\mathbf{y}|\mathbf{x}) = h(\mathbf{y})\phi(\mathbf{x}|\mathbf{y}) / f(\mathbf{x}) \quad (6)$$

Resulta evidente, que  $\phi$  y  $h$ , no se obtienen únicamente determinando  $f(x)$ , por lo tanto, debemos asumir supuestos acerca de su forma. El supuesto fundamental de los modelos de variables latentes es el de independencia condicional, el cual establece que dado los valores de las variables latentes, las variables manifiestas son independientes unas de otras. Este puede ser explicado como sigue:

$$\phi(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^p \phi_i(x_i|\mathbf{y}) = \phi_1(x_1|\mathbf{y}) \cdot \phi_2(x_2|\mathbf{y}) \cdot \dots \cdot \phi_p(x_p|\mathbf{y}) \quad (7)$$

El supuesto de probabilidad condicional implica que las variables latentes son las que producen las relaciones observadas en las variables manifiestas, en otras palabras, la interdependencia entre las variables manifiestas es debida a su dependencia común con las variables latentes, una vez que esta ha sido determinada el comportamiento de las variables manifiestas es esencialmente aleatorio. Otro supuesto adicional al de independencia condicional, es que por lo general se asume que las distribuciones de  $\phi_i$  y  $h$  son conocidas, pero dependen de un conjunto de valores de parámetros desconocidos (Everitt, 1984)

### 1.5.1 Estimación y ajuste en LVM

Asumiendo que el modelo de variables latentes está identificado, se procede a la estimación de parámetros y evaluación del ajuste.

Las variables manifiestas del modelo se pueden expresar mediante una matriz de covarianza poblacional, para ello, los elementos de la matriz son dados como una función particular de los parámetros del modelo  $\theta_1, \theta_2, \dots, \theta_t$ , donde  $t$  es el número de parámetros en el modelo.

Las estimaciones de los parámetros se obtienen minimizando mediante la función de ajuste la discrepancia entre  $\mathbf{S}$ , y  $\sum(\boldsymbol{\theta})$ , donde  $\mathbf{S}$  es la matriz de covarianza insesgada de una muestra de  $n$  observaciones sobre las  $p$  variables manifiestas y  $\sum(\boldsymbol{\theta})$  es la matriz de covarianza poblacional reproducida por el modelo y  $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_t)$

Varias funciones de discrepancia indican la magnitud de la diferencia entre los elementos de  $\mathbf{S}$  y las estimaciones propuestas en  $\sum(\boldsymbol{\theta})$ , las tres más comunes son las siguientes:

(1) Ordinary least squares (OLS):

$$F(\mathbf{S}, \sum(\boldsymbol{\theta})) = \sum_{i < j} (s_{ij} - \sigma_{ij}(\boldsymbol{\theta}))^2 \quad (8)$$

donde  $s_{ij}$  y  $\sigma_{ij}(\boldsymbol{\theta})$  son los elementos de  $\mathbf{S}$  y  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  respectivamente

(2) Generalized least squares (GLS):

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = \frac{1}{2} \text{traza}[(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{S}^{-1}]^2 \quad (9)$$

(3) Máxima Verosimilitud (ML),(asumiendo normalidad):

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = \log_e |\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \log_e |\mathbf{S}| + \text{traza}[\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}] - p \quad (10)$$

El primer método no es independiente de la escala de las variables manifiestas, en consecuencia, usando la matriz de covarianzas o correlaciones de la muestra puede producir valores diferentes de  $\boldsymbol{\theta}$ . En el segundo los elementos de  $\mathbf{S}$  generalmente están correlacionados y tienen varianzas diferentes, por lo que una simple medida de la diferencia entre  $\mathbf{S}$  y  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  puede no resultar apropiada. El tercer criterio se basa en la transformación del logaritmo de verosimilitud de las observaciones bajo el supuesto de que se distribuyen como una distribución normal multivariante (Everitt, 1984). El valor de  $\boldsymbol{\theta}$  que maximiza la función de verosimilitud se toma como el estimador de  $\boldsymbol{\theta}$ . El principio en el que se basa el método de máxima verosimilitud es el de obtener estimadores de los parámetros desconocidos, que maximicen la probabilidad de obtener dichas muestras (Santisteban, 1990)

Algunas propiedades de estas funciones de discrepancia son:

- a)  $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) \geq 0$
- b)  $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = 0$  si y sólo si  $\mathbf{S} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$
- c)  $F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$  es continua en  $\mathbf{S}$  y  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$

Browne (1982) muestra que si  $\boldsymbol{\theta}$  está identificado, todos los estimadores que se hallaron minimizando la función de discrepancia son consistentes. Si se

asume que las observaciones siguen una distribución normal multivariante para el GLS y ML, la función de discrepancia:  $(n-1)\min F(\mathbf{S}, \sum(\boldsymbol{\theta}))$  puede ser usada para probar el ajuste del modelo bajo la hipótesis de que la matriz de covarianza poblacional es de la forma que predice el modelo, este criterio tiene, una distribución chi-cuadrado asintótica con  $\nu = (p(p+1)-t)/2$  grados de libertad.

### 1.5.2 Máxima verosimilitud conjunta (JML)

Consiste en estimar simultáneamente los parámetros de los ítems y de los sujetos. La muestra de datos puede escribirse formalmente por una matriz  $\mathbf{x} = (x_{ij})$ , siendo  $i=1, \dots, I$  un subíndice sobre los sujetos y  $j=1, \dots, J$  un subíndice sobre los ítems. Cada respuesta  $x_{ij}$  de  $\mathbf{x}$  indica la respuesta del sujeto  $i$  al ítem  $j$ . Los parámetros de los sujetos se indican por el vector  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_I)$  y el vector  $\boldsymbol{\varepsilon}_j$  contiene los parámetros del ítem  $j$ . El objetivo del JML es obtener los estimadores  $\hat{\boldsymbol{\theta}}$  y  $\hat{\boldsymbol{\varepsilon}}$  de los vectores  $\boldsymbol{\theta}$  y  $\boldsymbol{\varepsilon}$  a partir de la matriz  $\mathbf{x}$ , (Abad et al., 2006).

El logaritmo de la función de verosimilitud es:

$$l(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \sum_{i=1}^I \sum_{j=1}^J \log P(x_{ij}) \quad (11)$$

### 1.5.3 Máxima verosimilitud marginal (MML):

El procedimiento MML se aplica sobre la tabla de contingencia de los patrones de respuesta. Si se considera el tamaño muestra  $I$  como una constante, existen  $H=2^J$  patrones de respuestas, la frecuencia de cada patrón de respuesta es  $n_h$  y, por tanto, la proporción es  $p_h = n_h/I$ . La distribución de la tabla de contingencia es multinomial, (Abad et al., 2006):

$$S(x) = \frac{I!}{\prod_{h=1}^H n_h!} \prod_{h=1}^H \pi_h^{n_h} \quad (12)$$

Donde  $\pi_h$  es la probabilidad de cada patrón y su estimador  $\hat{\pi}_h$  es  $p_h$ . Bajo el supuesto de que  $f(\theta)$  sigue una distribución normal o normal estándar, la probabilidad de tomar un sujeto al azar de la distribución  $f(\theta)$  y que este sujeto de cómo respuesta el patrón de respuesta  $h$  es:

$$\prod_{j=1}^J P(x_{hj}) f(\theta) \quad (13)$$

Como  $\theta$  no puede observarse, en la estimación de los parámetros de los ítems se utiliza la probabilidad marginal del patrón de respuestas  $\tau_h$ :

$$\tau_h = \int_{-\infty}^{+\infty} \prod_{j=1}^J P(x_{hj}) f(\theta) d\theta \quad (14)$$

Utilizando la probabilidad  $\tau_h$  en lugar de  $\pi_h$ , la distribución multinomial es:

$$L(x) = \frac{I!}{\prod_{h=1}^H n_h!} \prod_{h=1}^H \tau_h^{n_h} \quad (15)$$

El método MML consiste en maximizar la función  $L(x)$  con respecto a  $\epsilon$ , por sencillez se utiliza el algoritmo:

$$\log L(x) = K + \sum_{h=1}^H n_h \log \int_{-\infty}^{+\infty} \prod_{j=1}^J P(x_{hj}) f(\theta) d\theta \quad (16)$$

Donde  $K$  es una constante procedente del coeficiente multinomial y no tiene efecto en la estimación, (Abad et al., 2006).

#### 1.5.4 Máxima verosimilitud condicional (CML)

Este procedimiento se basa en la existencia de estadísticos suficientes para los parámetros del modelo, que son estadísticos que contienen toda la información de una muestra acerca de un determinado parámetro. En este procedimiento, en vez de la función de verosimilitud de un determinado patrón de respuestas, condicionada a un cierto nivel de aptitud, lo que se calcula es dicha función condicionada a una determinada puntuación total (Alvarado & Santisteban, 2001). Se suele aplicar a modelos de la familia de Rasch para estimar los parámetros de los ítems, aunque podría aplicarse también a los de los sujetos, una descripción más detallada de este procedimiento puede verse en Abad et al, 2006. Para una comparación entre CML y MML puede verse Thissen (1982)

### **1.5.5 Máxima verosimilitud información completa (FIML)**

Generalmente los métodos para estimar parámetros de variables ordinales, como el análisis factorial y los SEM, se basan en el uso de información univariada y bivariada y asumen el supuesto de normalidad. El método FIML (Bock & Aitkin, 1981) difiere según la función de respuesta al ítem adoptada, por ejemplo Muraki (1990) y Moustaki (2003) usan la función de respuesta logística y Muraki & Carlson (1995) usan la función de ojiva normal, ambas aproximaciones están basadas en el algoritmo EM, el cual es principalmente usado para estimar parámetros por ML con datos incompletos o valores perdidos (Skron dal & Rabe-Hesketh, 2004). Bajo el enfoque FIML se puede organizar la información de diferentes maneras esto le permite al investigador analizar con detalles los datos y tomar decisiones sobre los procedimientos a seguir, basado en el número posible de patrones de respuesta, la frecuencia de estos y el *Coverage Ratio* (CR), para una detalle de esta organización ver Jöreskog & Moustaki (2006).

En la estimación FIML se considera que la probabilidad incondicional  $\pi_r$  de un patrón de respuestas es:

$$\pi_r(\boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \pi_r(\xi) h(\xi) d\xi \quad (17)$$

Donde  $\int$  es una integral k-dimensional que se evalúa por la cuadratura de Gauss-Hermite. La probabilidad  $\pi_r$  es una función del vector de parámetros  $\boldsymbol{\theta}$  que contiene todos los interceptos  $\alpha_s^{(i)}$  y todas las cargas factoriales  $\beta_{ij}$  (en terminología del análisis factorial). Si  $n_r$  es la frecuencia con que ocurre el patrón de respuestas  $\mathbf{x}_r$  y  $p_r = n_r/N$  donde  $N$  es el tamaño de la muestra (la correspondiente probabilidad  $\pi_r$  fue definida en la ecuación 17), el logaritmo de la función de verosimilitud es:

$$\ln L = \sum_r n_r \ln \pi_r(\boldsymbol{\theta}) = N \sum_r p_r \ln \pi_r(\boldsymbol{\theta}) \quad (18)$$

Jöreskog & Moustaki (2001) con el objeto de maximizar ecuación 18 plantean la conveniencia de minimizar la función de ajuste:

$$F(\boldsymbol{\theta}) = \sum_r p_r [\ln p_r - \ln \pi_r(\boldsymbol{\theta})] = \sum_r p_r \ln [p_r / \pi_r(\boldsymbol{\theta})] \quad (19)$$

Esta función es no negativa e igual a cero sólo cuando el ajuste es perfecto, es decir, cuando  $p_r = \pi_r$  para todo  $r$ . El valor mínimo de  $F$  lo proporciona el estadístico  $\chi_{LR}^2$  (likelihood ratio chi-square) y es usado para probar el modelo.

La diferencia principal entre MML y FIML se basa en que mientras MML maximiza la verosimilitud de los datos condicionada sobre el rasgo latente, FIML maximiza la verosimilitud incondicional. La ventaja de MML comparada con FIML es que éste demanda menos coste computacional

## 1.6 Modelos de rasgo latente (LTM): Aproximación UV



Al igual que en el AF los LTM puede ser tratados desde la aproximación UV o desde la IRT. Los LTM son considerados como un tipo de LVM en los que las variables observadas son categóricas y las variables latentes son continuas (Bartholomew & Knott, 1999). Si las variables de respuesta  $x_i$  son ordinales o nominales, la probabilidad de cada patrón de respuesta, como una función de variables latentes  $z_1, z_2, \dots, z_k$  viene dada por:

$$\Pr(x_1 = a_1, x_2 = a_2, \dots, x_p = a_p | z_1, z_2, \dots, z_k) = f(z_1, z_2, \dots, z_k) \quad (20)$$

Donde  $a_1, a_2, \dots, a_p$  representan las diferentes categorías de respuestas de  $x_1, x_2, \dots, x_p$  respectivamente.

Como se dijo anteriormente, en la aproximación UV las variables binarias son asumidas como variables continuas ficticias, las cuales no deben confundirse con variables observadas ni con variables latentes. También podrían denominarse: variables incompleta observadas (Bartholomew et al., 2002).

Al igual que en el AF para datos ordinales, se asume que, para cada variable  $x_i$  hay una variable incompleta observada (pseudocontinua o continua ficticia)  $x_i^*$  que se distribuye normalmente con media  $\mu_i$  y varianza  $\sigma_i^2$ . La conexión entre  $x_i$  y  $x_i^*$  es la siguiente: cuando la UV  $x_i^*$  toma valores por debajo del umbral  $\tau_i$ , el ítem binario  $x_i$  toma el valor 1, en otro caso,  $x_i$  toma el valor 0. Los  $\tau_i$  son llamados parámetros de umbral.

La esencia del método consiste en que las  $x_i^*$ s han sido generadas bajo el modelo de análisis factorial clásico y esto supone que:

$$x_i^* = \alpha_{i0}^* + \alpha_{i1}^* y_1 + \alpha_{i2}^* y_2 + \dots + \alpha_{iq}^* y_q + e_i \quad (i = 1, \dots, p) \quad (21)$$

Donde  $\alpha_{ij}^*$  ( $j=1, \dots, q$ ) son las cargas factoriales, las  $y_j$ s son las variables latentes y  $e_i$  ( $i=1, \dots, p$ ) son los errores.

## 1.7 Modelos de Rasgo latentes: aproximación IRT

### 1.7.1 Respuestas binarias

Considerando que  $(y_1, \dots, y_p)$  son las  $p$  variables observadas u indicadores, Moustaki & Knott (2005) indican que para datos binarios se ajusta el modelo logístico a la probabilidad de obtener una respuesta positiva/correcta (respuesta 1) condicionada sobre las variables latentes  $\mathbf{z} = (z_1, \dots, z_q)$ . El modelo es

$$\text{logit } \pi_i(\mathbf{z}) = \alpha_{i0} + \sum_{j=1}^q \alpha_{ij} z_j \quad (22)$$

donde  $\pi_i(\mathbf{z}) = P(y_i = 1 | \mathbf{z})$

La probabilidad condicional de la variable  $y_i$  dado el vector de variables latentes  $\mathbf{z}$  es:

$$g(y_i | \mathbf{z}) = \pi_i(\mathbf{z})^{y_i} (1 - \pi_i(\mathbf{z}))^{1-y_i} \quad (23)$$

El parámetro  $\alpha_{i0}$  es usado para calcular la probabilidad de que un individuo medio ( $\mathbf{z} = 0$ ) responda positivamente al ítem  $i$  desde:

$$\pi_i(0) = \frac{e^{\alpha_{i0}}}{1 + e^{\alpha_{i0}}}$$

las cargas factoriales  $\alpha_{ij}$  muestran el efecto de las variables latentes sobre el logit de la probabilidad de una respuesta correcta. Para el caso de una variable latente el

modelo es conocido como modelo logístico de dos parámetros (Bock & Aitkin, 1981)

### 1.7.2 Respuestas ordinales:

Si la variable observada  $y_i$  es del tipo Likert con  $C_i$  categorías de respuesta, el modelo usado frecuentemente en aplicaciones es el proporcional odds models (POM) (Agresti, 1990). En este modelo el log de las probabilidades de que la distribución acumulada,  $\gamma_{is}(\mathbf{z})$ , sea igual o se encuentre debajo de una categoría de respuesta  $s$  es una función de las variables latentes  $\mathbf{z}$ . El modelo es:

$$\text{logit } \gamma_{is}(\mathbf{z}) = \alpha_{i0(s)} - \sum_{j=1}^q \alpha_{ij} z_j \quad (24)$$

Donde  $\gamma_{is}(\mathbf{z}) = P(y_i \leq s | \mathbf{z})$  y  $s = 1, \dots, C_i - 1$

La probabilidad condicional de la variable  $y_i$  dado el vector de variables latentes  $\mathbf{z}$  es:

$$g(y_i | \mathbf{z}) = \prod_{s=1}^{C_i} (\pi_i(\mathbf{z}))^{y_i(s)} \quad (25)$$

donde  $y_i(s)$  toma el valor 1 si el indicador  $i$  tiene la categoría  $s$  y 0 en cualquier otro caso.

Los parámetros  $\alpha_{i0(s)}$  son usualmente llamados puntos de cortes o umbrales y los  $\alpha_{ij}$  son llamadas cargas factoriales. Como se observa este modelo coincide con el explicado en el análisis factorial desde la IRT en el caso de modelos politómicos (1.4.1).

## 1.8 Conclusiones del capítulo

En el presente capítulo se da una breve exposición acerca de los LVM, su clasificación y su relación con los datos ordinales. Se hizo un especial énfasis en el modelo general de variables latentes que se origina a partir del AF que también se explica con detalle. Se muestran con detalles las relaciones entre los modelos lineales y los modelos IRT. Se repasa brevemente la esencia de los métodos de estimación de parámetros y su relación con la evaluación del ajuste del modelo. Además se introducen las aproximaciones teóricas para el tratamiento de datos binarios y ordinales desde las aproximaciones UV e IRT. En el siguiente capítulo se detallará en las bases de estas aproximaciones, haciendo un énfasis complementario en la estimación y el ajuste de modelos de rasgo latente desde estas aproximaciones teóricas.

## Capítulo II

# Tratamiento de datos ordinales en modelos de rasgo latente (LTM)

### Resumen:

En el presente capítulo se lleva a cabo un estudio comparativo que muestra las diferencias en la evaluación del ajuste bajo distintas aproximaciones teóricas. Desde el enfoque de las underlying variables (UV) se emplean métodos de estimación de información limitada GLS y WLS, desde la teoría de respuesta al ítem (IRT) el ajuste se hace con las funciones logística (POM) y normal (NOR), con métodos de estimación de información completa. Inicialmente se ajustan los datos empíricos con cada método de estimación a una estructura bifactorial del constructo analizado y estos resultados se toman como parámetros para el diseño de la simulación Monte Carlo. Los resultados muestran que el método de estimación adoptado afecta la bondad de ajuste del modelo. Los análisis permiten concluir que la aproximación IRT ofrece mejores ajustes que la aproximación UV, especialmente cuando se emplea el método POM.

## 2.1 Introducción

El tratamiento de datos ordinales in modelos de rasgo latente (LTM) se realiza generalmente desde dos aproximaciones, desde su conexión con los modelos de ecuaciones estructurales (SEM), considerando las denominadas underlying variables (UV) y la aproximación desde la teoría de respuesta al ítem (IRT), (Jöreskog & Moustaki, 2001; Bartholomew, et al. 2002; Moustaki, 2003).

En la aproximación UV se asume que las variables son ficticias continuas, pues no son variables observadas, ni variables latentes (Bartholomew, et al., 2002). Los métodos de estimación bajo este enfoque son de información limitada y utilizan la matriz de correlaciones policóricas (PM) o la matriz de covarianzas asintóticas (ACM) como entrada de datos para la estimación de los parámetros. Los procedimientos de estimación que se utilizan suelen ser los propuestos por Muthén (1984) y Jöreskog (1990, 1994), que son muy similares. Ambos proponen que la estimación se realice en tres etapas: en la primera, se estiman los estadísticos de primer orden (umbrales, medias y varianzas) por máxima verosimilitud (ML). En la segunda, Muthén propone estimar los estadísticos de segundo orden (correlaciones policóricas) por ML condicional, en tanto que Jöreskog estima además la ACM. La tercera etapa es la de estimación de los parámetros del modelo, que son los pesos factoriales: Muthén adopta el método de los mínimos cuadrados generalizados (GLS) y Jöreskog el de los mínimos cuadrados ponderados (WLS), siendo ACM la matriz de pesos.

El GLS a pesar de asumir normalidad multivariante (D'Agostino, 1986), se utiliza con datos no normales, usando la matriz de correlaciones policóricas (Jöreskog, 1990). También, en condiciones de no normalidad, se emplean métodos basados en la distribución libre de los datos (ADF), (Browne, 1984), como el WLS, aunque este presenta limitaciones en cuanto al tamaño de la muestra y al número de variables. Métodos alternativos, aunque bastante similares son, entre otros, los propuestos por Lee, Poon, & Bentler (1990, 1992) y Arminger & Küsters (1988).

Los fundamentos de la aproximación IRT se encuentran en los trabajos de Samejima (1969, 1997), Muraki (1990) y de Muraki & Carlson (1995). Los métodos que se emplean son de información completa y adoptan como funciones de enlace las distribuciones logística (proporcional odds models, POM) y normal (NOR). Una ventaja de estos procedimientos es que se pueden asumir distribuciones de los datos distintas a la normal y que la unidad de análisis son los patrones de respuesta de los sujetos, evitando así la pérdida de información (Moustaki, 2000). Por otra parte, los modelos IRT, son menos restrictivos en cuanto al número de variables usadas en el modelo, sin embargo, al ajustar modelos con un elevado número de factores, el costo computacional es elevado (Bock, et al. 1988). Los modelos IRT, así como otros de variables latentes, son tratados con detalle en Bartholomew, et al. (2002) y más recientemente por Ackerman, (2005).

## 2.2 Aproximaciones teóricas

### 2.2.1 Aproximación UV

El ajuste de cualquier función es un caso especial del ajuste de funciones de estructura de covarianza que puede venir dada por la ecuación (Browne, 1984, Jöreskog, 1990):

$$F(\boldsymbol{\theta}) = (\mathbf{s} - \boldsymbol{\sigma})' \mathbf{W}^{-1} (\mathbf{s} - \boldsymbol{\sigma}) = \sum_{g=1}^k \sum_{h=1}^g \sum_{i=1}^k \sum_{j=1}^i w^{gh,ij} (s_{gh} - \sigma_{gh})(s_{ij} - \sigma_{ij}) \quad (1)$$

Donde  $\mathbf{s}' = (s_{11}, s_{21}, s_{22}, s_{31}, \dots, s_{kk})$  es el vector de covarianzas de la matriz  $\mathbf{S}$  de orden  $k \times k$  usada para ajustar el modelo a los datos y  $\boldsymbol{\sigma}' = (\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \dots, \sigma_{kk})$  es el vector de elementos de la matriz  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  que reproduce los parámetros  $\boldsymbol{\theta}$  del modelo y  $w^{gh,ij}$  es un elemento típico de  $\mathbf{W}^{-1}$ , que es una matriz definida positiva

de orden  $u \times u$ , donde  $u = \frac{1}{2}k(k+1)$ . Para estimar los parámetros  $\theta$  del modelo, la función  $F(\theta)$  se minimiza con respecto a  $\theta$  (Jöreskog, 1990).

En el ajuste por el método GLS, cada elemento  $w^{gh,ij}$  en la Ecuación 1 viene expresado por:

$$w^{gh,ij} = N(2 - s_{gh})(2 - s_{ij})(s^{gi}s^{hi} + s^{gj}s^{hi}) \quad (2)$$

La forma usual de escoger  $\mathbf{W}$  en WLS es hacer que  $w^{gh,ij}$  sea una estimación consistente de la covarianza asintótica entre  $s_{gh}$  y  $s_{ij}$ . Si este es el caso, se dice que  $\mathbf{W}^{-1}$  es la matriz de pesos corregida.

En la teoría clásica de estructura de covarianzas (Browne, 1974; Jöreskog, 1981) se asume que las varianzas y covarianzas asintóticas de los elementos de  $\mathbf{S}$  son de la forma:

$$ACov(s_{hg}, s_{ij}) = (1/N)(\sigma_{gi}\sigma_{hj} + \sigma_{gh}\sigma_{hi}) \quad (3)$$

En el método WLS, cada elemento  $w^{gh,ij}$  en la Ecuación 1 viene expresado por:

$$w^{gh,ij} = m_{ghij} - s_{gh}s_{ij} \quad (4)$$

Donde  $m_{ghij} = \left(\frac{1}{N}\right) \sum_{a=1}^N (z_{ag} - \bar{z}_g)(z_{ah} - \bar{z}_h)(z_{ai} - \bar{z}_i)(z_{aj} - \bar{z}_j)$  son los momentos de cuarto orden.

### 2.2.2 Aproximación IRT



Una forma general de modelos lineales para variables ordinales la presenta McCullagh (1980) donde:

$$Link[\gamma_{is}(\mathbf{z})] = \alpha_{is} - \sum_{j=1}^k \beta_{ij} z_j \quad i = 1, 2, \dots, p;$$

representa la probabilidad de respuesta acumulada en la categoría  $s$ , o la probabilidad acumulada bajo el ítem  $x_i$ . Bajo el supuesto de que cada patrón de respuesta es independiente (independencia condicional), la probabilidad acumulada de una respuesta del ítem  $x_i$  a la categoría  $s$  o inferior puede escribirse como:

$$\gamma_{is}(\mathbf{z}) = \Pr(x_i \leq s) = \pi_{i1}(\mathbf{z}) + \pi_{i2}(\mathbf{z}) + \dots + \pi_{is}(\mathbf{z}) \quad (5)$$

La función de enlace es una función monótonamente creciente y puede ser logit, la función log-log complementaria, la función normal, la normal inversa o la función inversa de Cauchy.

Jöreskog & Moustaki (2001); Moustaki & Knott (2000) y Jöreskog & Moustaki (2006), consideran los casos particulares cuando la función de enlace es normal (NOR) y logística (POM) partiendo de que  $x_i$  es una variable ordinal con  $m_i \geq 2$  categorías ordenadas  $a_i = 1, 2, \dots, m_i$ , el modelo definido por la función de respuesta acumulada:

$$\Pr\{x_i \leq s | \mathbf{z}\} = F\left(\alpha_s^{(i)} - \sum_{j=1}^k \beta_{ij} z_j\right)$$

donde  $F$  es una función de distribución.

La función NOR que especifica que la probabilidad condicional de una respuesta del ítem  $x_i$  a la categoría  $s$  o inferior es:

$$\gamma_s^{(i)}(\mathbf{z}) = \int_{-\infty}^{\left[\alpha_s^{(i)} - \beta_{i1}Z_1 - \beta_{i2}Z_2 \dots \beta_{ik}Z_k\right]} \phi(u) du \quad (6)$$

Donde  $\Phi(u) = \left[ \alpha_s^{(i)} - \sum_{j=1}^k \beta_{ij} z_j \right]$  es la función de distribución normal estándar, (Jöreskog & Moustaki, 2001). Los  $\alpha_s^{(i)}$  son los parámetros interceptos. Hay un intercepto para cada variable y para cada categoría. Para definir la ordinalidad los parámetros interceptos deben satisfacer:

$$\alpha_1^{(i)} < \alpha_2^{(i)} \dots < \alpha_{m_i-1}^{(i)} < \alpha_{m_i}^{(i)}$$

Los parámetros  $\beta_{ij}$  son las cargas factoriales.  $\alpha_s^{(i)}$  y  $\beta_{ij}$  se conocen como parámetros no estandarizados. Similarmente podemos obtener la probabilidad condicional de  $\pi_s^{(i)}(\mathbf{z})$  para la categoría  $a$  de la variable  $i$ :

$$\pi_a^{(i)}(\mathbf{z}) = \Phi \left[ \alpha_a^{(i)} - \sum_{j=1}^k \beta_{ij} z_j \right] - \Phi \left[ \alpha_{a-1}^{(i)} - \sum_{j=1}^k \beta_{ij} z_j \right]$$

La ecuación anterior se conoce como: función de categoría de respuesta (category response function).

Si  $\mathbf{x}_r = (x_1 = a_1, x_2 = a_2, \dots, x_p = a_p)$  es un patrón de respuesta dado. Bajo el supuesto de independencia condicional, la probabilidad condicional  $\pi_r$  del patrón de respuesta  $\mathbf{x}_r$  dado un  $\mathbf{z}$  es:

$$\pi_r(\mathbf{z}) = \prod_{i=1}^p \pi_{ai}^{(i)}(\mathbf{z}) = \prod_{i=1}^p \left[ \gamma_{ai}^{(i)}(\mathbf{z}) - \gamma_{ai-1}^{(i)}(\mathbf{z}) \right] \quad (7)$$

la cual se obtiene integrando  $\pi_r(\mathbf{z})$  en un espacio k-dimensional:

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \pi_r(\mathbf{z}) h(\mathbf{z}) d(\mathbf{z}) \quad (8)$$

Donde  $h(\mathbf{z})$  es la función de densidad de  $\mathbf{z}$ , y se resuelve por la cuadratura de Gauss-Hermite:

$$\pi_r = \sum_{t_1=1}^{v_1} \dots \sum_{t_k=1}^{v_k} \pi_r(z_{1t_1}, \dots, z_{kt_k}) h(z_{1t_1}) \dots h(z_{kt_k}) \quad (9)$$

Donde  $z_{1t_1}, \dots, z_{kt_k}$  son los nodos de la cuadratura y  $h(z_{1t_1}), \dots, h(z_{kt_k})$  son los correspondientes pesos expresados mediante polinomios Hermite de grado menor y  $v_1, v_2, \dots, v_k$  son el número de cuadraturas a lo largo de cada dimensión.

La función POM supone que las  $m_i$  categorías de respuesta ordenada de la variable  $x_i$  tienen probabilidades condicionales independientes  $(\pi_1^{(i)}(\mathbf{z}), \pi_2^{(i)}(\mathbf{z}), \dots, \pi_{m_i}^{(i)}(\mathbf{z}))$  que son una función de un vector de orden  $k$  de variables latentes  $\mathbf{z}$ . La función POM se corresponde con el modelo logístico:

$$\ln \left[ \frac{\gamma_s^{(i)}(\mathbf{z})}{1 - \gamma_s^{(i)}(\mathbf{z})} \right] = \alpha_s^{(i)} - \sum_{j=1}^k \beta_{ij} z_j$$

Donde  $s = 1, \dots, m_{i-1}$  y

$$\gamma_s^{(i)}(\mathbf{z}) = \Pr(x_i \leq s) = \pi_1^{(i)}(\mathbf{z}) + \dots + \pi_s^{(i)}(\mathbf{z})$$

es la probabilidad de respuesta en la categoría  $s$  o inferior de la variable  $i$ . Hay un parámetro intercepto  $\alpha_s^{(i)}$  para cada categoría, mientras que el parámetro  $\beta_{ij}$ , que representa las cargas factoriales, es el mismo entre las categorías de la misma variable. De igual manera que con la función NOR los parámetros interceptos  $\alpha_s^{(i)}$  deben satisfacer:

$$\alpha_1^{(i)} < \alpha_2^{(i)} \dots < \alpha_{m_{i-1}}^{(i)} < \alpha_{m_i}^{(i)} = \infty$$

la función POM, puede escribirse como:

$$\gamma_s^{(i)}(\mathbf{z}) = \frac{\exp \left[ \alpha_s^{(i)} - \sum_{j=1}^k \beta_{ij} z_j \right]}{1 + \exp \left[ \alpha_s^{(i)} - \sum_{j=1}^k \beta_{ij} z_j \right]} = \Psi \left[ \alpha_s^{(i)} - \sum_{j=1}^k \beta_{ij} z_j \right] \quad (10)$$

Donde  $\Psi(x)$  es la función de distribución de la función logística.

La probabilidad condicional del patrón de respuesta  $\mathbf{x}_r$  dado un  $\mathbf{z}$  es:

$$\pi_r(\mathbf{z}) = \prod_{i=1}^p \pi_{ai}^{(i)}(\mathbf{z}) = \prod_{i=1}^p \left[ \gamma_{ai}^{(i)}(\mathbf{z}) - \gamma_{ai-1}^{(i)}(\mathbf{z}) \right] \quad (11)$$

Al igual que en la función NOR se obtiene integrando  $\pi_r(\mathbf{z})$  en un espacio  $k$ -dimensional:

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \pi_r(\mathbf{z}) h(\mathbf{z}) d(\mathbf{z})$$

Donde  $h(\mathbf{z})$  es la función de densidad de  $\mathbf{z}$ , y se resuelve por la cuadratura de Gauss-Hermite. Otros detalles relativos a los fundamentos de la función POM pueden verse en Skrondal & Rabe-Hesketh (2004).

### 2.2.3 Estandarización de parámetros

Los parámetros  $\alpha_a^{(i)}$  y  $\beta_{ij}$  en los modelos anteriores no se encuentran estandarizados. Para fines interpretativos es conveniente considerar los parámetros estandarizados  $\tau_a^{(i)}$  y  $\lambda_{ij}$ , Bartholomew & Knott (1999); Joreskog & Moustaki (2001), los definen como:

$$\tau_a^{(i)} = \frac{\alpha_a^{(i)}}{\left(1 + \sum_{j=1}^k \beta_{ij}^2\right)^{\frac{1}{2}}} \text{ y } \lambda_{ij} = \frac{\beta_{ij}}{\left(1 + \sum_{j=1}^k \beta_{ij}^2\right)^{\frac{1}{2}}} \quad (12)$$

Los parámetros estandarizados se corresponden uno a uno y se pueden obtener también en forma viceversa, como sigue:

$$\alpha_a^{(i)} = \frac{\tau_a^{(i)}}{\left(1 - \sum_{j=1}^k \lambda_{ij}^2\right)^{\frac{1}{2}}} \text{ y } \beta_{ij} = \frac{\lambda_{ij}}{\left(1 - \sum_{j=1}^k \lambda_{ij}^2\right)^{\frac{1}{2}}} \quad (13)$$

De manera intuitiva  $\tau_a^{(i)}$  y  $\lambda_{ij}$  representan los umbrales y las cargas factoriales en el enfoque UV,  $\alpha_a^{(i)}$  y  $\beta_{ij}$  representan los parámetros de los ítem en modelos IRT de dos parámetros.

#### 2.2.4 Ajuste del modelo

Para la evaluación del ajuste con los métodos GLS y WLS se utilizan los índices globales:  $\chi^2$ ,  $\chi^2/df$ , RMSEA, GFI, AGFI y también el CFI que tradicionalmente se usa en los SEM, (Bollen, 1989; Bollen & Long, 1993; Arbuckle, 1999; Loehlin, 2004). Bajo la aproximación IRT, cuando el número de variables manifiestas ordinales es grande, se espera que muchos patrones de respuesta tengan una frecuencia esperada menor que 5 y muchos serán tan pequeños que no ocurrirán (Moustaki, 2003). En estos casos, es aconsejable calcular el estadístico  $\chi^2$  likelihood-ratio ( $\chi^2_{LR}$ ) o el  $\chi^2$  global-fit ( $\chi^2_{GF}$ ) por pares o tríos de variables, donde:

$$\chi^2_{LR} = 2 \sum_r n_r \ln(p_r / \hat{\pi}_r) = 2N \sum_r p_r \ln(p_r / \hat{\pi}_r) = 2NF(\hat{\theta}) \quad (14)$$

$$\chi^2_{GF} = \sum \left[ (n_r - N\pi_r)^2 / (N\pi_r) \right] = N \sum (p_r - \hat{\pi}_r)^2 / \hat{\pi}_r \quad (15)$$

con grados de libertad igual al número de patrones de respuestas diferentes menos uno menos el número de elementos independientes de  $\theta$ .

Cada celda sigue una distribución  $\chi^2$  con 1 grado de libertad, entonces los residuales mayores que 4 indican un pobre ajuste al nivel de significación  $\alpha = 0,05$  (Jöreskog & Moustaki, 2001). La evaluación del ajuste basado en el análisis de información bivariada considera cada suma de  $\chi^2_{LR}$  y  $\chi^2_{GF}$ , llamadas

$S_{LR}$  and  $S_{GF}$ , respectivamente. Las reglas de interpretación pueden resumirse como: (a) si cualquiera de los  $S_{LR}$  o  $S_{GF}$  es mayor que  $4m_i m_j$  (donde  $m_i$  son las  $m$  categorías de la variable  $i$  y  $m_j$  son las  $j$  categorías de la variable  $j$  en la tabla de contingencia) el ajuste es pobre, aunque también se ha propuesto un umbral más bajo en  $3m_i m_j$ ; (b) si cualquiera de los  $S_{LR}$  o  $S_{GF}$  es mayor que el valor de  $\chi^2_{[(mixmj)-1]}$  a un nivel de significación  $\alpha = 0.01$ , se considera un mal ajuste (Bartholomew et al, 2002).

## 2.3 Comparación entre métodos de estimación

### 2.3.1 Problemas resueltos y no resueltos

La comparación entre métodos de estimación en las aproximaciones UV e IRT también ha sido objeto de estudio. Jöreskog & Moustaki (2001) utilizando datos simulados comparan métodos de estimación de información limitada como el underlying bivariate normal (UBN) con los basados en información completa como el underlying multivariate normal (UMN), POM y NOR. Maydeu-Olivares (2005), con el propósito de establecer un marco teórico unificado, compara la estimación de parámetros modelos lineales y no lineales de la IRT con el análisis factorial, usando ML y unweighted least squares (ULS) con información limitada y con información completa. Moustaki et al. (2004) comparan los métodos POM y NOR de información completa con robust maximum likelihood y con el WLS de información limitada, haciendo énfasis el efecto covariable sobre las variables latentes. Stephenson & Holbert (2003) comparan distintas aproximaciones de modelos de variables latentes con FIML, similarmente Stapleton (2006) compara métodos de estimación en SEM. Todos estos estudios con variables ordinales se centran en la comparación de los parámetros, estudiando la precisión de las estimaciones y apenas abordan los efectos que tiene el método de estimación empleado en el ajuste de los modelos.

### 2.3.2 Una solución aproximada:

El objetivo del presente capítulo es mostrar cómo el ajuste con datos ordinales se ve afectado por el método de estimación elegido (GLS, WLS, NOR y POM) y determinar cuál es la mejor opción para ajustar estos modelos LTM. La validación de los índices de ajuste obtenidos se realiza a través de un procedimiento de simulación Monte Carlo.

## **2.4 Método:**

Para ilustrar los procedimientos de estimación y ajuste, hemos utilizado los datos obtenidos de la aplicación de una escala de acción política, cuyos ítems y propiedades psicométricas son descritos en la sección del instrumento. Los valores estimados son usados como parámetros para el subsiguiente proceso de simulación, el cual es usado para validar los índices de ajuste obtenidos para el modelo de dos factores.

### **2.4.1 Aproximación empírica:**

#### **2.4.1.1 Sujetos**

Un total de 643 participantes (media de edad 28 años, SD=10.2), el 47% hombres y 53% mujeres, cumplieron una escala para la medida de la acción política. Los valores perdidos fueron excluidos por “listwise deletion”, es decir, que aquellos patrones de respuesta que tuvieron un valor perdido en alguno de los ítems, fueron omitidos del análisis, reduciéndose la muestra a 564 participantes.

#### **2.4.1.2 Instrumento**

Para este trabajo usamos una modificación de la escala de acción política, que mide acciones políticas convencionales, del ISSP Citizenship (CIS, 2004; Carton et al., 2004; Wilson et al., 2005). La versión que se utiliza cambia un ítem de la escala original (GRUPO) y añade dos ítems más (ILEGAL y VIOLEN) que

conforman un segundo factor, acciones políticas no convencionales (Brady, 1999; Anduiza & Bosch, 2004). Así la escala usada contiene 10 ítems con respuestas tipo Likert de 5 puntos con las categorías: (1) Ni participó, ni quizás lo haga (NN), (2) No ha participado, pero podría hacerlo (NP), (3) Alguna vez ha participado (AP), (4) Participó durante el año pasado (PP), (5) Participó recientemente (PR).

Las propiedades psicométricas de los componentes simples y la puntuación total de la escala (en particular, la consistencia interna y la validez de constructo) fueron determinadas y probadas satisfactoriamente. La consistencia interna de la escala global es  $\alpha = 0.78$  y la de los factores son  $\alpha = 0.78$  and  $\alpha = 0.74$  para acciones políticas convencionales y no convencionales respectivamente. Esos valores son similares a los obtenidos por la aplicación del ISSP (2004) en diferentes países: Australia  $\alpha = 0.77$ ; Inglaterra  $\alpha = 0.79$ ; Francia  $\alpha = 0.73$ ; España  $\alpha = 0.79$ ; E.E.U.U  $\alpha = 0.81$  y Venezuela  $\alpha = 0.69$ . Un análisis factorial confirmatorio ha mostrado un buen ajuste de los datos al modelo de dos factores ( $\chi^2/df = (93.16/34) = 2.74$ ; GFI = 0.99; AGFI = 0.98; CFI = 0.95 y RMSEA = 0.056).

Los 10 ítems son:

- Firmar una petición (FIRMAR)
- Comprando o dejando de comprar ciertos productos por razones políticas, éticas o para favorecer el medio ambiente (BOICOT)
- Asistir a una reunión o mitin político (ASIREU)
- Contactar o intentar contactar con un político o funcionario para expresarle sus opiniones (CONPOL)
- Entregar dinero o contactar fondos para una actividad social o política (ENTDIN)
- Contactar o comparecer ante los medios de comunicación para expresar sus opiniones (CONMED)
- Colaborar con un grupo o plataforma de acción ciudadana (COLCIU)



- Participar actividades ilegales de protesta (cortar tráfico, invasiones, etc.) (LEGAL)
- Desarrollar algún tipo de acción violenta (VIOLEN)
- Participar en un foro o grupo de discusión política en Internet (INTERN)

#### 2.4.1.3 Software

El ajuste de los modelos se ha realizado con LISREL 8.8 (Jöreskog & Sörbom, 2006a). En aproximación UV se ha utilizado el modulo PRELIS 2.81 (Jöreskog & Sörbom, 2006b) para la generación de las matrices PM y ACM. Para la estimación de los parámetros con métodos GLS y WLS se ha utilizado LISREL 8.8. Para la aproximación IRT se ha utilizado la versión 2.81 del PRELIS mediante el comando de ordinal factor análisis (OFA) con las funciones NOR y POM

#### 2.4.1.4 Estimación y ajuste de los datos empíricos

##### Solución de un factor:

Empezaremos el análisis ajustando el modelo de un factor a las 10 variables. Los resultados del análisis se basan el modelo de la IRT usando como funciones de enlace las funciones probit (NOR) y logit (POM) (Jöreskog & Moustaki 2001). También se evaluará la aproximación UV, bajo los enfoques de Muthen (1984) y Jöreskog (1994) generando matrices policóricas y la matriz de covarianza asintótica con los métodos de estimación GLS y WLS. Todos los ajustes de los modelos se evalúan con LISREL.

Las cargas factoriales estandarizadas y la estimación de los errores estándar para las dos aproximaciones se muestran en la tabla 2.1. Las estimaciones POM son generalmente mayores que las NOR. De igual manera, las estimaciones WLS son sustancialmente mayores que las GLS. También se observa que las estimaciones son más próximas entre NOR y las soluciones bajo

UV, particularmente con el GLS, esto se debe a que la aproximación NOR asume normalidad al igual que los modelos bajo GLS, en consecuencia esto no corresponde a una diferencia entre modelos sino a métodos de estimación (Moustaki, 2004). La tabla 2.1 también muestra que errores estándar en general son bajos.

Tabla 2.1  
Cargas estandarizadas y errores para la solución de un factor

ÍTEM	IRT		UV	
	NOR	POM	GLS	WLS
FIRMAR	0.530 (0.042)	0.743 (0.042)	0.42 (0.05)	0.58 (0.03)
BOICOT	0.425 (0.044)	0.647 (0.050)	0.35 (0.05)	0.51 (0.04)
ASIREU	0.701 (0.031)	0.865 (0.031)	0.58 (0.04)	0.70 (0.03)
CONPOL	0.756 (0.034)	0.902 (0.011)	0.64 (0.04)	0.77 (0.02)
ENTDIN	0.629 (0.035)	0.818 (0.034)	0.55 (0.04)	0.68 (0.03)
CONMED	0.652 (0.035)	0.844 (0.024)	0.60 (0.04)	0.67 (0.03)
COLCIU	0.641 (0.035)	0.832 (0.015)	0.58 (0.04)	0.61 (0.03)
ILEGAL	0.434 (0.071)	0.622 (0.077)	0.71 (0.05)	0.83 (0.03)
VIOLEN	0.367 (0.090)	0.558 (0.112)	0.67 (0.05)	0.79 (0.04)
INTERN	0.604 (0.037)	0.801 (0.028)	0.60 (0.05)	0.59 (0.03)

Notas: IRT: Ítem response Theory, NOR: Normal Model POM: Proportional Odds ratios, UV: Underlying Variable, GLS: Generalized Least Square, WLS: Weighted Least Square

Para la evaluación del ajuste bajo el enfoque IRT se utiliza el análisis bivariado del conjunto de ítems analizando la suma de cada uno de los  $\chi^2_{LR}$  y  $\chi^2_{GF}$ , llamadas  $S_{LR}$  y  $S_{GF}$ , respectivamente, siguiendo la regla de interpretación de Bartholomew, et al, 2002. Para la aproximación UV se utilizan los índices globales de ajuste:  $\chi^2/df$ , CFI, GFI, AGFI y RMSEA (Bollen, 1989; Arbuckle, 1999; Loehlin, 2004).

Tabla 2.2

Matriz de sumas de valores  $\chi^2$  por pares de variables. Solución NOR de un factor.

(a) Índice  $\chi^2_{LR}$

	FIRMAR	BOICOT	ASIREU	CONPOL	ENTDIN	CONMED	COLCIU	ILEGAL	VIOLEN	INTERN
FIRMAR	0,155									
BOICOT	50.198	0,041								
ASIREU	48.091	41.126	0,169							
CONPOL	28.957	49.435	40.826	0,133						
ENTDIN	29.085	27.353	38.852	33.601	0,222					
CONMED	24.972	42.442	21.949	66.080	33.433	0,07				
COLCIU	30.203	39.199	46.481	52.853	27.085	72.506	0,254			
ILEGAL	21.830	21.190	18.549	24.234	11.891	32.328	44.066	0,002		
VIOLEN	26.228	14.837	24.376	38.723	22.694	48.561	37.715	98.669	0,037	
INTERN	47.327	25.900	34.777	29.480	37.360	36.008	47.938	38.176	42.158	0,029

Total univariate (LR<sub>uni</sub>) = 1.112  
 Total bivariate (LR<sub>biv</sub>) = 1669.741

(b) Índice  $\chi^2_{GF}$

	FIRMAR	BOICOT	ASIREU	CONPOL	ENTDIN	CONMED	COLCIU	ILEGAL	VIOLEN	INTERN
FIRMAR	0,155									
BOICOT	52.892	0,041								
ASIREU	49.629	46.332	0,169							
CONPOL	31.277	54.149	42.671	0,133						
ENTDIN	30.175	27.969	39.680	34.791	0,221					
CONMED	28.450	46.601	23.821	70.872	35.430	0,07				
COLCIU	31.880	43.563	45.905	59.663	27.696	78.627	0,254			
ILEGAL	21.000	18.097	16.965	23.303	10.301	36.945	49.442	0,002		
VIOLEN	25.065	13.141	21.160	41.682	19.448	55.863	40.422	172.032	0,038	
INTERN	49.669	26.828	34.739	30.066	37.719	38.267	64.122	41.877	46.889	0,029

Total univariate (GF<sub>uni</sub>) = 1.111  
 Total bivariate (GF<sub>biv</sub>) = 1837.114

Las tablas 2.2 y 2.3, muestran respectivamente, el análisis univariado y multivariado de  $\chi^2_{LR}$  y  $\chi^2_{GF}$ , para las soluciones NOR y POM. En ella se puede observar, que en ambos métodos (NOR y POM) se muestra un pobre ajuste a la solución de un factor.

Tabla 2.3

Matriz de sumas de valores  $\chi^2$  por pares de variables. Solución POM de un factor.

(a) Índice  $\chi_{LR}^2$

	FIRMAR	BOICOT	ASIREU	CONPOL	ENTDIN	CONMED	COLCIU	ILEGAL	VIOLEN	INTERN
FIRMAR	0,101									
BOICOT	43.708	0,016								
ASIREU	41.900	37.414	0,068							
CONPOL	23.480	46.366	35.229	0,162						
ENTDIN	31.381	24.475	38.171	32.930	0,405					
CONMED	22.006	38.864	20.584	57.163	31.835	0,038				
COLCIU	26.901	33.636	44.062	50.065	26.324	66.516	0,32			
ILEGAL	21.029	21.062	17.184	22.674	12.525	29.866	44.348	0,012		
VIOLEN	25.240	14.106	23.389	38.236	23.771	50.048	39.787	104.718	0,026	
INTERN	45.663	22.226	32.781	25.745	34.795	30.007	44.414	39.031	43.718	0,004
Total univariate (LR <sub>uni</sub> ) = 1.152										
Total bivariate (LR <sub>biv</sub> ) = 1579.371										

(b) Índice  $\chi_{GF}^2$

	FIRMAR	BOICOT	ASIREU	CONPOL	ENTDIN	CONMED	COLCIU	ILEGAL	VIOLEN	INTERN
FIRMAR	0,101									
BOICOT	45.619	0,016								
ASIREU	42.555	41.058	0,068							
CONPOL	25.040	49.150	36.297	0,161						
ENTDIN	32.832	24.459	39.245	33.413	0,399					
CONMED	24.294	41.736	22.093	58.332	33.697	0,038				
COLCIU	27.796	36.550	42.296	54.332	26.170	70.715	0,318			
ILEGAL	18.797	18.625	15.376	20.548	11.000	30.254	44.773	0,012		
VIOLEN	21.871	11.866	20.408	39.516	20.087	56.534	43.649	202.256	0,026	
INTERN	48.047	22.785	31.293	25.240	33.949	30.927	56.344	45.746	48.469	0,004
Total univariate (GF <sub>uni</sub> ) = 1.143										
Total bivariate (GF <sub>biv</sub> ) = 1726.038										

El valor de  $S_{LR}$  en el índice  $\chi_{LR}^2$  en el método NOR de 98, 669 (Tabla 2.2a), es superior al umbral de 75 (3x5x5), de igual manera para el índice  $\chi_{GF}^2$  (Tabla 2.2b), los  $S_{GF}$  alcanzan valores de 78,627 y 172,032 que superan el umbral establecido. Por otra parte los valores  $S_{LR}$  y  $S_{GF}$  para la función POM, en ambos índices, exceden claramente el umbral 75,  $S_{LR}$  con 104,718 (Tabla 2.3a) y

$S_{GF}$  con 202,256 (Tabla 2.3b), lo que confirma el mal ajuste de la aproximación IRT para la solución de un factor.

La tabla 2.4 muestra los índices de ajuste global para la solución bajo el enfoque UV, en general revelan el pésimo ajuste de la aproximación de un factor bajo la aproximación UV. Los peores índices de ajuste, se observan en la solución UV GLS.

Tabla 2.4  
Aproximación UV: índices de ajuste para la solución de un factor

	$\chi^2/df$	GFI	AFI	CFI	RMSEA
GLS	9,44	0.54	0.28	0.004	0.16
WLS	6,76	0.96	0.94	0.82	0.10

### Solución de dos factores:

La tabla 2.5 muestra las cargas factoriales estandarizadas y la estimación de los errores estándar para cada método de estimación para la solución de dos factores.

Tabla 2.5  
Cargas factoriales estandarizadas y errores para la solución de dos factores

#### a) Aproximación IRT

Ítem	NOR		POM	
	F1	F2	F1	F2
FIRMAR	0.540 (0.042)		0,743 (0,035)	
BOICOT	0.407 (0.046)		0,620 (0,053)	
ASIREU	0.706 (0.035)		0,860 (0,019)	
CONPOL	0.752 (0.031)		0,894 (0,031)	
ENTDIN	0.595 (0.042)		0,781 (0,031)	
CONMED	0.604 (0.040)		0,780 (0,044)	
COLCIU	0.592 (0.039)		0,769 (0,044)	
ILEGAL		0.674 (0.048)		0,853 (0,052)
VIOLEN		0.970 (0.200)		0,977 (0,020)
INTERN	0.522 (0.048)		0,678 (0,061)	

## b) Aproximación UV

Ítem	GLS		WLS	
	F1	F2	F1	F2
FIRMAR	0.53 (0.04)		0.64 (0.03)	
BOICOT	0.42 (0.05)		0.48 (0.04)	
ASIREU	0.72 (0.03)		0.76 (0.03)	
CONPOL	0.79 (0.03)		0.83 (0.02)	
ENTDIN	0.62 (0.04)		0.67 (0.03)	
CONMED	0.64 (0.04)		0.69 (0.03)	
COLCIU	0.63 (0.04)		0.70 (0.03)	
ILEGAL		0.88 (0.08)		0.94 (0.07)
VIOLEN		0.72 (0.08)		0.78 (0.07)
INTERN	0.60 (0.04)		0.62 (0.03)	

Notas: IRT: Item response theory; NOR: normal model; POM: proportional odds ratios; UV: underlying variable; GLS: generalized least squares; WLS: weighted least squares.

Como se observa las estimaciones POM son generalmente mayores que las NOR (Tabla 2.5a) y las WLS son sustancialmente mayores que las GLS (Tabla 2.5b). También se observa que las estimaciones son más próximas entre NOR y las soluciones bajo UV, particularmente con el GLS. Esto se debe a que la aproximación NOR asume normalidad al igual que los modelos bajo GLS, en consecuencia, esto no se debe a la diferencia entre los modelos, sino a los métodos de estimación (Moustaki, et al. 2004). La solución WLS, muestra cargas factoriales estandarizadas superiores a GLS y NOR, pero no a POM. La solución WLS sólo supera a la POM en el valor de la carga factorial de la variable ILEGAL del segundo factor. Los errores estandarizados son pequeños y similares en las aproximaciones UV e IRT.

Tabla 2.6

Matriz de sumas de valores  $\chi^2$  por pares de variables. Solución NOR de dos factores.

(a) Índice  $\chi^2_{LR}$

	FIRMAR	BOICOT	ASIREU	CONPOL	ENTDIN	CONMED	COLCIU	ILEGAL	VIOLEN	INTERN
FIRMAR	0.165									
BOICOT	50.869	0.077								
ASIREU	47.427	41.043	0.08							
CONPOL	29.295	49.584	39.216	0.239						
ENTDIN	28.730	27.533	38.750	33.913	0.248					
CONMED	24.121	42.241	21.339	66.876	33.783	0.122				
COLCIU	29.349	39.551	46.494	52.403	27.673	72.887	0.235			
ILEGAL	21.248	21.253	17.325	21.249	12.271	32.616	44.363	0.345		
VIOLEN	23.491	13.877	24.691	34.758	19.167	38.421	34.269	21.568	1.309	
INTERN	48.477	26.451	35.395	28.549	36.788	36.178	48.234	36.706	36.760	0.151

Total univariate ( $LR_{uni}$ ) = 2.971

Total bivariate ( $LR_{biv}$ ) = 1557.183

(b) Índice  $\chi^2_{GF}$

	FIRMAR	BOICOT	ASIREU	CONPOL	ENTDIN	CONMED	COLCIU	ILEGAL	VIOLEN	INTERN
FIRMAR	0.166									
BOICOT	53.871	0.078								
ASIREU	48.939	46.564	0.080							
CONPOL	32.223	55.100	41.785	0.243						
ENTDIN	30.356	28.401	39.919	35.785	0.253					
CONMED	27.260	46.968	23.181	72.429	36.689	0.123				
COLCIU	30.605	44.131	46.067	59.635	28.931	80.373	0.237			
ILEGAL	21.055	19.791	16.878	20.886	11.528	42.546	55.249	0.354		
VIOLEN	26.607	13.706	25.477	38.957	18.031	46.087	39.721	25.800	1.377	
INTERN	50.657	27.695	35.686	29.186	37.048	39.023	65.783	41.926	40.680	0.151

Total univariate ( $GF_{uni}$ ) = 3.061

Total bivariate ( $GF_{biv}$ ) = 1699.219

Table 2.7

Matriz de sumas de valores  $\chi^2$  por pares de variables. Solución POM de dos factores.

(a) Índice  $\chi^2_{LR}$

	FIRMAR	BOICOT	ASIREU	CONPOL	ENTDIN	CONMED	COLCIU	ILEGAL	VIOLEN	INTERN
FIRMAR	0.281									
BOICOT	45.735	0.181								
ASIREU	42.354	37.800	0.522							
CONPOL	24.398	47.002	34.846	0.625						
ENTDIN	30.937	24.614	38.985	33.713	0.296					
CONMED	21.351	38.968	20.965	58.938	32.216	0.335				
COLCIU	26.150	34.351	45.030	50.034	27.103	67.185	0.311			
ILEGAL	21.310	20.998	17.094	20.688	11.849	30.357	43.047	0.154		
VIOLEN	27.946	19.218	26.345	40.414	25.826	49.769	40.605	27.760	6.079	
INTERN	47.626	23.126	33.983	25.489	34.460	30.492	44.981	35.598	42.652	0.126

Total univariate ( $LR_{uni}$ ) = 8.910

Total bivariate ( $LR_{biv}$ ) = 1524.310

(b) Índice  $\chi^2_{GF}$

	FIRMAR	BOICOT	ASIREU	CONPOL	ENTDIN	CONMED	COLCIU	ILEGAL	VIOLEN	INTERN
FIRMAR	0.283									
BOICOT	48.211	0.184								
ASIREU	43.299	42.546	0.531							
CONPOL	26.446	51.673	37.546	0.654						
ENTDIN	34.144	25.124	41.480	36.122	0.300					
CONMED	23.910	43.160	23.066	62.628	36.590	0.342				
COLCIU	26.699	37.637	43.952	55.726	28.511	73.850	0.312			
ILEGAL	19.947	19.298	16.020	19.323	10.637	35.993	47.680	0.155		
VIOLEN	27.954	19.133	28.195	50.658	25.564	66.223	49.631	36.060	6.326	
INTERN	49.717	23.842	32.972	25.209	33.651	32.716	57.884	36.669	51.549	0.126

Total univariate ( $GF_{uni}$ ) = 9.213

Total bivariate ( $GF_{biv}$ ) = 1658.850

Las tablas 2.6 y 2.7 muestran las matrices de valores  $S_{LR}$  y  $S_{GF}$  para los índices  $\chi^2_{LR}$  y  $\chi^2_{GF}$ , respectivamente, por pares de variables. En la Tabla 2.6a el índice  $\chi^2_{GF}$  presenta valores de  $S_{GF}$  inferiores a 75 ( $3m_i m_j$ ), lo que indicaría un



ajuste satisfactorio, sin embargo el índice  $\chi^2_{GF}$  (Tabla 2.6b) revela un valor de  $S_{GF}$  de 80,313, lo cual invalida la adecuación de la solución. Con el método POM, los índices  $\chi^2_{LR}$  (Tabla 2.7a) y  $\chi^2_{GF}$  (Tabla 2.7b) presentan en todos los casos valores de  $S_{LR}$  y  $S_{GF}$  inferiores a 75, lo que supone un buen ajuste del modelo a los datos con este método.

Para las soluciones UV se muestran índices de ajustes globales poco satisfactorios para la estimación por GLS, como es el caso del  $\chi^2/df > 3$ , sin embargo, los parámetros estimados bajo WLS muestran un buen ajuste  $\chi^2/df < 3$ , tal y como se observa en la Tabla 2.8.

Tabla 2.8  
Aproximación UV: índices de ajuste para la solución de dos factores

	$\chi^2/df$	GFI	AGFI	CFI	RMSEA
GLS	5.61	0.91	0.85	0.92	0.047
WLS	2.74	0.99	0.98	0.95	0.056

#### 2.4.2 Simulación Monte Carlo

El objetivo del procedimiento de simulación es generar un conjunto de soluciones ajustando modelos con cada método de estimación (GLS, WLS, POM y NOR) que permitan evaluar los índices:  $\chi^2/df$ , RMSEA, GFI, AGFI y CFI para los métodos GLS y WLS y la suma de los elementos  $S$  de la matriz para los índices  $\chi^2_{LR}$  y  $\chi^2_{GF}$  expresadas como LR-b y GF-b, para los métodos POM y NOR. Por lo tanto, se considera que la variable independiente es el método de estimación y la variable dependiente son los índices de ajuste. En la simulación Monte Carlo se utilizan como parámetros las estimaciones obtenidas con los datos empíricos para la generación de muestras de tamaño 600 de 10 variables ordinales no normales a través del PRELIS, usando la matriz de correlaciones policóricas y los valores de los umbrales. La generación de números aleatorios, en PRELIS, está basada en la distribución uniforme implementada en la función URAND que ejecuta un algoritmo en fortran desarrollado por Schrage (1979).

Este método, conocido como generadores de congruencia, es uno de los más utilizados para generar valores uniformes. Para la estimación de los parámetros los métodos GLS y WLS, utilizan las matrices PM y ACM para ajustar la estructura bifactorial con LISREL, mientras que en la aproximación IRT, para cada réplica se lleva a cabo un ordinal factor análisis mediante el comando OFA del PRELIS y las estimaciones se obtienen ajustando a las funciones NOR y POM. En total se realizan 800 réplicas, obteniéndose 200 soluciones para cada método de estimación. Antes de evaluar los resultados se eliminan las soluciones no-convergentes (según el criterio de LISREL tras 250 iteraciones)

#### 2.4.2.1 Resultados de la simulación

Analizando los resultados obtenidos para cada método de estimación se encontró inconsistencia en el ajuste con respecto al modelo evaluado con los datos empíricos. Los índices de ajuste para las muestra simuladas con los métodos GLS y WLS se muestran en la Tabla 2.9, donde se observa que los valores de los índices de ajuste incremental y global (GFI, AGFI y CFI) son bastante buenos, el RMSEA se mantiene en un rango aceptable, sin embargo, el valor del chi-cuadrado presenta valores altos en ambos métodos de estimación. Contrastando estos resultados con los de la Figura 2.1, se observan contradicciones con la solución obtenida con los datos empíricos, pues se obtiene estimaciones más bajas del  $\chi^2/df$  y del RMSEA (Figura 2.1a) bajo el método GLS, en comparación con las obtenidas por el método WLS.

En el caso de los métodos POM y NOR, los resultados son más estables y congruentes con la solución obtenida con los datos empíricos. Los estadísticos correspondientes a esta aproximación se encuentran en la Tabla 2.10 en donde se puede observar que hay muy poca dispersión de los valores de los índices LR-b y GF-b, lo que es un indicador de buen ajuste (Jöreskog & Moustaki, 2006).

Tabla 2.9  
Aproximación UV: estadísticos descriptivos para cada índice de ajuste obtenidos en la simulación Monte Carlo

	GLS				WLS			
	media	dt	sesgo	Kurtosis	media	dt	sesgo	kurtosis
$\chi^2$	117.846	17.377	0.227	0.395	127.408	17.762	0.078	0.590
RMSEA	0.062	0.007	-0.202	0.948	0.065	0.006	-0.336	1.066
GFI	0.986	0.002	-0.025	0.388	0.986	0.002	0.038	0.515
AGFI	0.978	0.003	-0.025	0.386	0.977	0.003	0.038	0.513
CFI	0.958	0.009	-0.020	0.650	0.963	0.007	-0.183	0.233

Tabla 2.10  
Aproximación IRT: estadísticos descriptivos para cada índice de ajuste obtenidos en la simulación Monte Carlo

	POM				NOR			
	media	dt	sesgo	kurtosis	media	dt	sesgo	kurtosis
LR <sub>biv</sub>	804.06	52.70	0.323	-0.382	821.11	71.70	0.858	1.40
GF <sub>biv</sub>	753.66	50.20	0.445	-0.158	788.87	79.20	0.940	1.27

Por otra parte, la figura 2.1b muestra que, bajo el método POM, las estimaciones son más bajas y por lo tanto mejores en comparación con el método NOR, de igual manera, se observa que bajo la solución POM apenas hay casos con valores extremos mientras que en las soluciones NOR los valores extremos son más frecuentes.

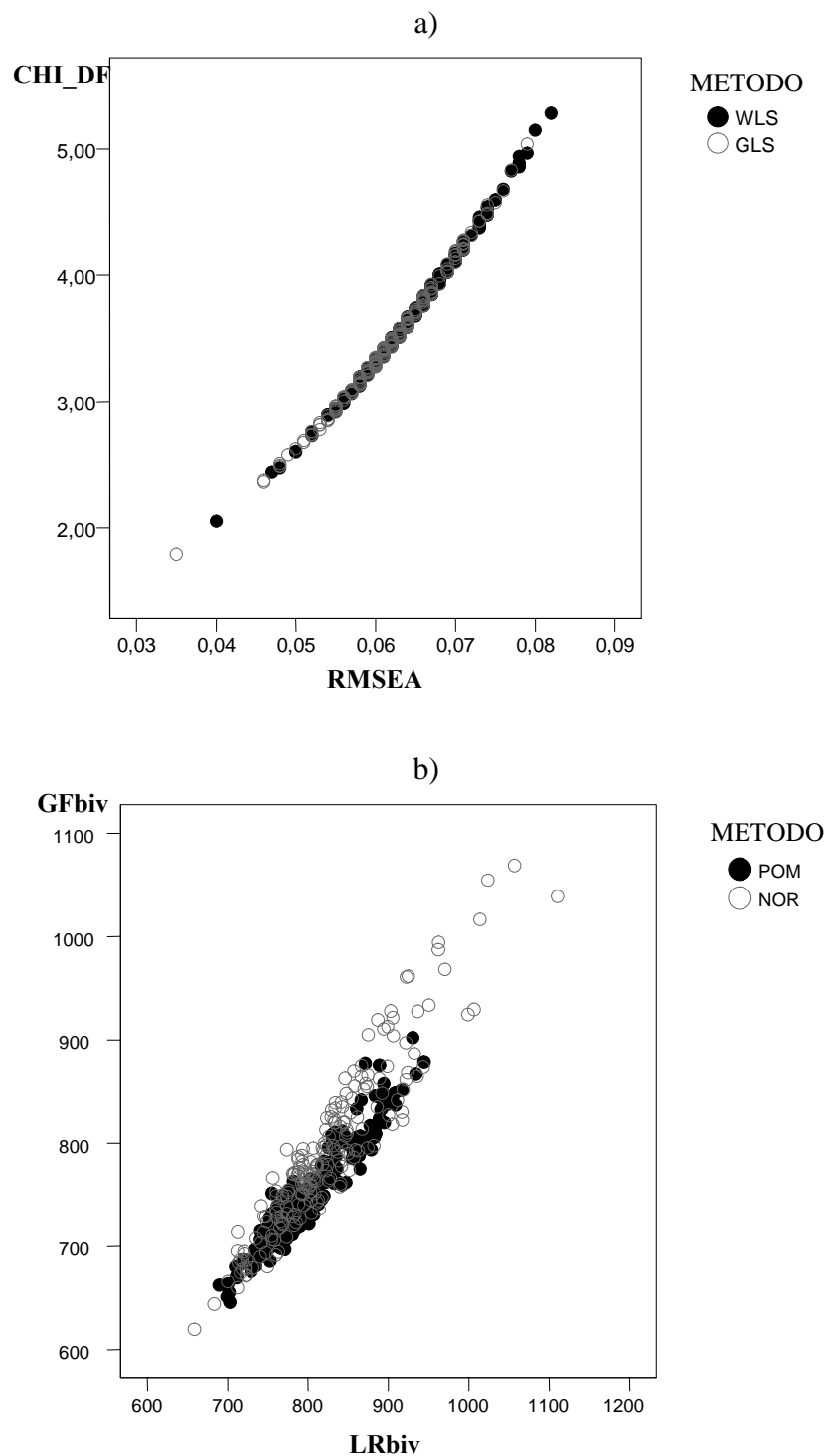


Figura 2.1  
Valores de los índices de ajuste por método de estimación: a) Aproximación UV;  
b) Aproximación IRT.

## 2.5 Discusión y conclusiones

Se ha evaluado la adecuación de distintos enfoques metodológicos (las aproximaciones UV e IRT) para el análisis de variables ordinales en LTM. La etapa inicial de estimación y de ajuste para cada uno de los cuatro métodos aplicados a datos empíricos permite concluir: a) las soluciones que muestran los mejores ajustes son las que no suponen normalidad multivariante WLS y POM, aunque POM proporciona mejores parámetros que WLS, b) en el enfoque UV la solución propuesta por Jöreskog (1994) de la utilización de la ACM estimada por WLS muestra un mejor ajuste que la propuesta de Muthén (1984) basada en el cálculo de la matriz de correlaciones policóricas y la posterior estimación de parámetros por GLS, c) en el enfoque IRT, la que muestra un mejor ajuste y, por lo tanto, la mejor solución, es la del modelo logístico con la función POM.

El procedimiento de simulación ha permitido validar los índices LR-b y GF-b obtenidos con los datos empíricos, pudiendo así comparar las soluciones que ofrecían mejor ajuste. En los datos simulados se mostró una mayor estabilidad y consistencia en las soluciones bajo la aproximación IRT, específicamente bajo la estimación POM, la cual muestra globalmente las mejores soluciones. Por otro lado, se revelan contrariedades e inconsistencias con respecto a los métodos WLS y GLS, los cuales en general no muestran buenos ajustes en las muestras simuladas.

La estimación por métodos de información completa (considerados en la aproximación IRT) muestra claras ventajas al producir estimaciones suficientes en comparación con el GLS y WLS, basados en la información limitada de las distribuciones marginales (aproximación UV). Esta ventaja no sólo se manifiesta en los ajustes de los modelos, sino también en las múltiples posibilidades de organización y evaluación de los datos, tal y cómo es referenciado por Jöreskog & Moustaki (2006).

Las reglas de interpretación del ajuste planteadas por Jöreskog & Moustaki (2001), e interpretadas por Bartholomew, et al. (2002), parecen adecuadas a las

diversas situaciones de análisis of ordinal data y se muestran compatibles con otras técnicas como la recategorización de las puntuaciones originales en MMRG, sin embargo, al interpretarlas, se puede caer en la arbitrariedad y resultan poco prácticas al evaluar un número elevado de soluciones, como sucede con las que se obtienen en un procedimiento de simulación.

La adecuación teórica del constructo analizado revela la calidad de los ajustes encontrados con el método POM, la cual se muestra en concordancia con la estructura bifactorial del constructo Acción Política en que uno de los factores lo componen acciones convencionales o acciones legales y el otro las acciones no convencionales o actividades ilegales de propuesta (Almond & Verba, 1970, Anduiza & Bosch, 2004, Brady, 1999).

Finalmente, esta investigación muestra que al probar el ajuste con datos ordinales, tanto los procedimientos GLS y WLS, aparecen como métodos adecuados, sin embargo, la simulación ha mostrado que sus soluciones no son estables, por lo que se recomienda el uso de los métodos IRT como una mejor alternativa para el ajuste de modelos de variables latentes con datos ordinales.

## Capítulo III

# Evaluación del ajuste y procedimientos de simulación in LTM

### Resumen

En el presente capítulo se busca indagar en los factores que afectan la evaluación de los índices de ajuste en LTM. A partir de los resultados encontrados en el capítulo anterior, que revelan que se obtienen diferentes valoraciones del ajuste, según el método de estimación adoptado, se considera una nueva variable al evaluar la precisión de los índices de ajuste: el procedimiento de simulación. Específicamente, se analiza el efecto de adoptar el Bootstrap (no paramétrico tal y como se encuentra disponible en el software comercial) o de técnicas de simulación como el Monte Carlo, al evaluar las distribuciones de los índices de ajustes. Este efecto es analizado conjuntamente con el método de estimación en LTM aproximados desde la IRT con métodos de estimación de información completa. Los resultados se obtienen a partir de un estudio de simulación considerando variables dependientes son medidas globales de ajuste en modelos LTM aproximados desde la IRT. Los resultados sugieren que el procedimiento de simulación afecta significativamente la evaluación del ajuste del modelo, estos hallazgos son contrastados con los que figuran en literatura sobre el tema.

### 3.1 Introducción

Diversos estudios de simulación han evaluado el comportamiento de los índices de ajuste y han determinado cómo estos índices son afectados por características tales como: el número de factores, el número de indicadores por factor y el tamaño de las cargas factoriales. Entre otras cosas, han demostrado que mientras la estimación de parámetros no se ve afectada por las características del modelo o el tamaño de la muestra, los índices de ajuste presentan importantes variaciones, tal es el caso, por ejemplo, del  $\chi^2$  para tamaños muestrales menores a 200 y de los índices GFI, AGFI y RMR que son atenuados por tamaños de muestras pequeños o por un número mayor de factores o de indicadores por factor (Anderson & Gerbing, 1984; Bearden, Sharma & Teel, 1982; Boomsma, 1982). El tamaño de la muestra y su efecto en los índices de ajuste ha sido estudiado en diversas investigaciones, bien sea evaluando su efecto con otras variables como la parsimonia del modelo (Marsh & Balla, 1994) o simplemente el efecto del tamaño de la muestra (Tanaka, 1987).

Las causas que afectan el ajuste, en modelos de variables latentes (LVM), han sido abordadas principalmente desde los modelos de ecuaciones estructurales, evaluando índices como:  $\chi^2$ , RMSEA, GFI, AGFI y CFI, entre otros, (Bollen, 1989; Bollen & Long, 1993; Arbuckle, 1999; Marsh, Hau & Graysin, 2005), pero escasamente han sido evaluados desde la teoría de respuesta al ítem (IRT), concretamente, no se ha evaluado el efecto de ciertas variables sobre índices de ajuste obtenidos a partir de la frecuencia observada y esperada de los patrones de respuestas en tablas de contingencia, como el  $\chi^2$  likelihood ratio ( $\chi^2_{LR}$ ) y el  $\chi^2$  global fits ( $\chi^2_{GF}$ ), o la suma de éstos  $S_{LR}$ , para el total de los  $\chi^2_{LR}$  y  $S_{GF}$  para el total de los  $\chi^2_{GF}$ , (Bartholomew, et al. 2002; Bartholomew & Knott, 1999; Jöreskog & Moustaki, 2006).

Los investigadores buscan constantemente índices de ajustes adecuados, y además, se proponen probar la calidad de los mismos resaltando su fiabilidad y validez. Para lograr este cometido, someten los valores de los índices obtenidos a



procesos de comprobación que se caractericen principalmente por su rigurosidad. Este proceso incluye técnicas de diversa naturaleza: tratamiento de datos perdidos, técnicas de simulación, métodos de estimación alternativos, entre otros. En cada una de estas técnicas se presentan diversas opciones, así como también la posibilidad de aplicar simultáneamente dos o más de ellas, la adecuada selección de los métodos y estrategias empleados se refleja en la precisión de los resultados y en la calidad del ajuste obtenido. Así por ejemplo, dentro de las técnicas de simulación para evaluar los índices de ajuste se puede optar por el Bootstrap (Bollen & Stine, 1993) o la simulación Monte Carlo (Gerbing & Anderson, 1993); en cuanto a la estimación se puede optar por métodos de información completa o limitada (Bolt, 2005); en el tratamiento de los datos perdidos se pueden adoptar diversas estrategias (Wiggins & Sacker, 2002; Davey, Savla & Luo, 2005).

En el capítulo II se demostró que el método de estimación de parámetros adoptado puede afectar los valores de los índices de ajuste, pues los resultados obtenidos al ajustar datos ordinales con métodos de estimación de información limitada como el GLS o WLS bajo la aproximación UV, no son los mismos que se obtienen bajo la aproximación IRT con métodos de estimación de información completa ajustando funciones normales o logísticas. Otras investigaciones, que han analizado el comportamiento de la bondad del ajuste con los métodos de estimación, también ponen de manifiesto tales relaciones, (Ximenez & García, 2005; Jöreskog & Moustaki, 2001; Moustaki, et al. 2004; Maydeu-O, 2005). Por otra parte, estudios han demostrado que no se obtiene la misma precisión en los estimadores implementado Bootstrap en lugar de Monte Carlo (Fan, 1994; Langeheine, et al. 1996).

### **3.2 Una solución aproximada**

En consecuencia, resulta de especial interés, estudiar el efecto conjunto que tienen ciertas variables sobre los índices de bondad de ajuste. En el presente capítulo se pretende evaluar el efecto que tiene la adopción de una determinada técnica de simulación para probar la precisión de los índices de ajuste,

considerando principalmente: el Bootstrap (Efron, 1979,1982, 1987, 1988; Efron & Tibshirani 1986, 1993) y Monte Carlo (Rubinstein, 1981; Mooney 1997; Skrondal 2000; Paxton, et al. 2001). A nivel técnico, la diferencia principal que se establece entre ambas técnicas es que en Bootstrap se toman muestras aleatorias de una muestra original de datos reales y en Monte Carlo las muestras aleatorias se pueden generar sin necesidad de disponer de una muestra original de datos (Jöreskog & Sörbom, 2002).

En este sentido se busca dar cuenta de los efectos que pueden tener las variables: procedimientos de simulación escogido para generar muestras (Bootstrap o Monte Carlo) y el método de estimación de información completa adoptado (POM o NOR) en los índices de ajuste de LVM aproximados desde la IRT. Para lograr este objetivo, se lleva a cabo un estudio de simulación Monte Carlo para examinar si el procedimiento de simulación empleado para la generar las muestras y el método de estimación afectan el ajuste de la solución expresada mediante los índices globales de ajuste.

### **3.3 Método:**

En primer lugar se hace breve exposición conceptual teórica, de las técnicas de simulación empleadas: Monte Carlo y Bootstrap, se prescinde de la explicación de los métodos de estimación, pues ya han sido explicados ampliamente en los capítulos anteriores. Posteriormente diseña el estudio de simulación, se evalúan los modelos mediante un ANOVA multivariante y se analizan sus resultados

#### **3.3.1 Bootstrap.**

La esencia del Bootstrap supone realizar una serie de cálculos intensivos repetitivos por medio de algoritmos para obtener con mayor precisión (*accuracy*) la medida de los estimadores (media, varianza, errores, entre otros) y la distribución muestral de estos. En el contexto inferencial esto supone una ventaja

frente a situaciones donde las asunciones distribucionales y formulas analíticas son inviables o insostenibles.

Una definición intuitiva del Bootstrap supone un conjunto de variables independientes representadas por vectores  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  tomadas de una muestra aleatoria de tamaño  $n$ . Sea  $F_n(\mathbf{X})$  la distribución de probabilidad empírica la cual asigna una probabilidad de  $1/n$  a cada punto  $x_1, x_2, \dots, x_n$ , cuyo estimador es  $\hat{\theta}$ . Si tomamos una muestra aleatoria de tamaño  $N < n$  con reposición obtenemos un nuevo estimador  $\hat{\theta}^*$ , si este proceso se repite  $B$  veces entonces obtenemos  $\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)$  estimadores. Al construir la distribución de probabilidad a partir de los  $B$   $\hat{\theta}^*(B)$  y asignando una probabilidad de  $1/B$  a cada punto  $\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)$ , obtenemos la función  $\hat{F}_n(\mathbf{X})$ , que corresponde a la distribución muestral de la estimación Bootstrap de  $\hat{\theta}$ . Esta distribución puede usarse para hacer inferencias sobre  $\theta$ .

Los pasos básicos para la estimación Bootstrap son los siguientes, Efron (1979); Hinckley (1988); Efron y Tibshirabi (1993):

1. Construir una distribución de probabilidad empírica,  $\hat{F}(X)$ , a partir de la muestra asignando una probabilidad de  $1/n$  a cada punto,  $x_1, x_2, \dots, x_n$ . Esta es la función de distribución empírica de  $X$ , la cual es el estimador no-paramétrico de máxima verosimilitud de la función de distribución de la población  $F(X)$ .
2. A partir de  $\hat{F}(X)$ , se extrae una muestra aleatoria simple de tamaño  $n$  con reposición, esta es una remuestra  $x_b^*$ .
3. Se calcula el estadístico de interés,  $\hat{\theta}$ , a partir de esa remuestra, obteniendo  $\hat{\theta}_b^*$ .
4. Se repiten los pasos 2 y 3  $B$  veces, donde  $B$  es un número grande. La magnitud de  $B$  en la práctica depende de las pruebas que se van a aplicar a los datos. En general,  $B$  debería ser de entre 50 a 200 para estimar el error típico de

$\hat{\theta}$ , y al menos de 1000 para estimar intervalos de confianza alrededor de  $\hat{\theta}$  (Efron y Tibshirani, 1986, 1993).

5. Construir una distribución de probabilidad a partir de los  $B$   $\hat{\theta}^*_b$  asignando una probabilidad de  $1/B$  a cada punto,  $\hat{\theta}^*_1, \hat{\theta}^*_2, \dots, \hat{\theta}^*_B$ . Esta distribución es la estimación Bootstrap de la distribución muestral de  $\hat{\theta}$ ,  $\hat{F}^*(\hat{\theta}^*)$  y puede usarse para hacer inferencias sobre  $\theta$ . El estimador Bootstrap del parámetro  $\theta$  se define como:

$$\hat{\theta}_{(.)}^* = \frac{\sum_{b=1}^B \hat{\theta}_b^*}{B}$$

es decir, como la media de los valores del estadístico calculados en las  $B$  remuestras bootstrap.

Efron (1979) indica que la dificultad del procedimiento Bootstrap es el cálculo de la distribución para sus estimadores y plantea tres métodos para la obtención de la misma: (i) cálculos directos, (ii) aproximación Monte Carlo y (iii) Series de Tylor extendidas. El procedimiento comúnmente aplicado es la simulación Monte Carlo.

Chernick, (1999) simplifica el procedimiento para obtener aproximaciones Bootstrap a partir de una aproximación por Monte Carlo:

1. Generar una muestra de tamaño  $n$  (donde  $n$  es el tamaño de la muestra original) con reemplazamiento de una distribución empírica (muestra Bootstrap).
2. Calcular  $\theta^*$ , el valor de  $\hat{\theta}$  es obtenido usando la muestra Bootstrap en lugar de la muestra original.
3. Repetir los pasos 1 y 2  $k$  veces

Posterior a la introducción del concepto de Bootstrap, esta técnica ha sido ampliamente adoptada para evaluar la precisión de los estimadores e índices de ajuste. Al mismo tiempo se han realizado nuevos aportes, incluso por el mismo autor (Efron, 1990), y se han planteado su aplicación con modificaciones por otros autores como Bollen & Stine (1993). Algunos autores han estudiado las potencialidades y fallas de este procedimiento, varios de ellos presentan

argumentos que cuestionan los resultados de su aplicación (Chernick, 1999, Young, 1994 y Fan, 1994). Algunos procedimientos Bootstrap para evaluación de parámetros pueden encontrarse en Bollen & Stine (1990, 1993); Craig (1992); Chan & Chan (2004) y para la evaluación del ajuste puede verse Bone, Sharma & Shimp (1989).

### **3.3.2 Monte Carlo**

Otra técnica disponible y ampliamente aplicada para simular muestras a partir de ciertos parámetros encontrados en investigaciones previas es Monte Carlo, este procedimiento permite considerar la naturaleza y distribución de los datos, y a partir de éstas, genera variables (continuas, discretas u ordinales) con distribuciones normales o distintas a ésta.

El algoritmo original de simulación Monte Carlo está fundamentado en la generación de números aleatorios por el método de transformación inversa, el cual se basa en las distribuciones acumuladas de frecuencias de variables aleatorias y en la generación de valores asociados a éstas mediante la distribución uniforme.

El principio sobre el que se sustenta el Monte Carlo es que el comportamiento de un estadístico en muestras aleatorias puede ser evaluado mediante un proceso empírico en el que se van extrayendo muestras y se evalúa su comportamiento. La estrategia para hacer esto es crear un mundo artificial, o pseudo-población, el cual se parezca en los aspectos más relevantes al mundo real. Esta pseudo-población consiste en un procedimiento matemático para generar muestras con un conjunto de números a partir de datos tomados de una población real (Mooney, 1997)

Mooney (1997), describe el procedimiento básico de Monte Carlo, como sigue:

1. Especificar la pseudo-población en términos simbólicos, como una forma que puede ser usada para generar muestras. Se puede desarrollar un algoritmo que permita generar datos con ciertas distribuciones.
2. La muestra de la pseudo-población (o pseudo.muestra) debe reflejar la forma del estadístico de interés.

3. Calcular  $\hat{\theta}$  en la pseudo-población y almacenarlo en un vector,  $\hat{\theta}$ .
4. Repetir los pasos 2 y 3  $t$  veces, donde  $t$  es el número de réplicas.
5. Construir una distribución de frecuencia relativa con el resultado de los  $t$  valores de  $\hat{\theta}_t$ , la cual es la distribución Monte Carlo estimada de la distribución muestral de  $\hat{\theta}$  bajo las condiciones especificadas por la pseudo-población y por el procedimiento de muestreo.

Robert & Casella (2004), tratan con detalle los diferentes algoritmos, métodos y aplicaciones de Monte Carlo

### 3.3.3 Simulación Monte Carlo para datos ordinales

Jóreskog & Sörbom, (2002) presentan mediante 5 pasos la generación de variables ordinales mediante Monte Carlo. Sean  $x$  e  $y$  dos variables ordinales con  $x$  e  $y$  categorías respectivamente. Un estudio Monte Carlo incluye 5 pasos:

1.- Escoger la correlación  $\rho$  en la población y los umbrales  $\alpha_1, \alpha_2, \dots, \alpha_{r-1}$  para las  $x$  y  $\beta_1, \beta_2, \dots, \beta_{s-1}$  ( $\alpha_0 = \beta_0 = -\infty; \alpha_r = \beta_s = +\infty$ ). Calcular las probabilidades  $\pi_{ij} = \Pr(x = i, y = j)$  como:

$$\pi_{ij} = \int_{\alpha_{i-1}}^{\alpha_i} \int_{\beta_{j-1}}^{\beta_j} \phi_2(u, v) du dv$$

donde  $\phi_2$  es la función de densidad de la normal estándar bivariada con correlación  $\rho$ .

2.- Generar una observación aleatoria  $x = i, y = j$  con probabilidad  $\pi_{ij}$ .

Repetir esto  $N$  veces. El resultado es una tabla de contingencia:

$$\begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1s} \\ n_{21} & n_{22} & \cdots & n_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ n_{r1} & n_{r2} & \cdots & n_{rs} \end{bmatrix}$$

3.- Calcular uno de los seis tipos siguientes de correlaciones, para variables continuas: producto momento de Pearson (raw scores), rangos de

Spearman, Coeficientes Tau b de Kendall, canónicas y para variables ordinales: producto momento de Pearson (normal scores) y policóricas.

4.- Repetir los pasos 2 y 3  $k$  veces.

5.- Calcular media, varianza, asimetría y el error medio cuadrático.

Algunas aplicaciones Monte Carlo pueden verse, por ejemplo, en Meade & Lautenschlager, 2004; Ximenez & García (2005) y Beaducel & Wittmann (2005).

### 3.4 Estudio de simulación

En el procedimiento de simulación Monte Carlo se siguieron las recomendaciones planteadas en Psychometrika, 1979; Skrondal, 2000. En primer lugar se formuló su objetivo principal: examinar si, en LTM ajustados mediante métodos de estimación de la información completa aproximados desde la IRT, el procedimiento de simulación seleccionado para la generar las muestras y el método de estimación afectan el ajuste de la solución expresada mediante las medidas globales de ajuste. Para ello, se consideran dos variables independientes: procedimiento de simulación (1.Bootstrap; 2.Monte Carlo) y método de estimación de información completa (1. NOR; 2. POM). La variable dependiente son las medidas de ajuste global en sus versiones univariadas y bivariadas (1.  $LR_{uni}$ ; 2.  $LR_{biv}$ ; 3.  $GF_{uni}$ ; 4.  $GF_{biv}$ ) que se definen más adelante y a los cuales nos referiremos como medidas globales de ajuste. Se incluyen todas las combinaciones posibles de los tratamientos. Se trata entonces de un diseño factorial 2x2 La solución original define dos estructuras factoriales teóricas y está conformada por diez variables observadas (Tabla 3.1)

Se establecen dos procedimientos de simulación para la generación de muestras, Bootstrap: tomando de los datos de la solución original una fracción de muestreo del 95% y Monte Carlo: a partir de la matriz de correlaciones policóricas y de los valores de los umbrales de la solución original, se generan una muestra de tamaño 600 de 10 variables ordinales no normales a través del PRELIS. La generación de números aleatorios está basada en la distribución

uniforme implementada en la función URAND del programa la cual implementa un algoritmo en fortran desarrollado por Schrage (1979), este método es conocido como generadores de congruencia son los más utilizados para generar valores uniformes (Revuelta & Ponsoda, 2003). El PRELIS adopta métodos para la generación de datos no normales basado en los trabajos de Fleishman (1978) y Vale & Maurelli (1983).

Para cada réplica se lleva a cabo un análisis factorial ordinal mediante el comando OFA del Prelis 2.81 (Jöreskog & Sorbom, 2006) y se obtienen la estimación de parámetros bajo métodos de estimación NOR y POM. La simulación y la estimación de parámetros se repiten para 200 veces para cada situación experimental de las variables independientes ( $200 \times 2 \times 2$ ) para un total de 800 réplicas ( $r=1, \dots, 800$ )

Tabla 3.1.  
Parámetros estimados e índices de ajustes modelo original de dos factores por método de estimación

	POM	NOR
$\lambda_1$	0.74	0.54
$\lambda_2$	0.62	0.41
$\lambda_3$	0.86	0.71
$\lambda_4$	0.89	0.75
$\lambda_5$	0.78	0.60
$\lambda_6$	0.78	0.60
$\lambda_7$	0.77	0.59
$\lambda_8 (f2)$	0.85	0.67
$\lambda_9 (f2)$	0.98	0.97
$\lambda_{10}$	0.68	0.52
LR-uni	8.910	2.971
LR-biv	1524.31	1557.18
GF-uni	9.213	3.061
GF-biv	1658.85	1699.22

Para el análisis de los resultados, se verifican inicialmente la convergencia de las soluciones, las soluciones no-convergentes son eliminadas (según el criterio



de LISREL tras 250 iteraciones). La variable dependiente está conformada por las medidas de ajuste proporcionadas por el programa PRELIS, que se obtienen como el total de los  $S_{LR}$  para los  $\chi^2_{LR}$  y el total de los  $S_{GF}$  para los  $\chi^2_{GF}$ , que en sus versiones univariadas y bivariadas se formulan como sigue:

$$LR_{uni} = 2N \sum_{i=1}^n p_{ii}^{(gh)} \ln \left( p_{ii}^{(gh)} / \hat{\pi}_{ii}^{(gh)} \right) \quad (1)$$

$$LR_{biv} = \sum_{i=1}^{\frac{1}{2}n(n-1)} F_{(LR)}^{(gh)}(\hat{\theta}) = 2N \sum_{i=1}^{\frac{1}{2}n(n-1)} \left[ \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} p_{ab}^{(gh)} \ln \left( p_{ab}^{(gh)} / \hat{\pi}_{ab}^{(gh)} \right) \right] \quad (2)$$

$$GF_{uni} = \sum_{i=1}^n [(p_{ii}^{(gh)} - \hat{\pi}_{ii}^{(gh)})^2 / (\hat{\pi}_{ii}^{(gh)})] \quad (3)$$

$$GF_{biv} = \sum_{i=1}^{\frac{1}{2}n(n-1)} F_{(GF)}^{(gh)}(\hat{\theta}) = \sum_{i=1}^{\frac{1}{2}n(n-1)} \left[ \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} (p_{ab}^{(gh)} - \hat{\pi}_{ab}^{(gh)})^2 / (\hat{\pi}_{ab}^{(gh)}) \right] \quad (4)$$

Siguiendo la recomendación de Skrandall (2000), se utiliza un modelo sencillo para analizar los resultados, que sólo incluye el efecto principal de cada factor y la doble interacción sobre la variable dependiente, el modelo es el siguiente:

$$\text{Indice: Intercept} + \text{Simul} + \text{Estima} + \text{Simul} * \text{Estima} \quad (5)$$

Donde: *Índice* son los índices globales de ajuste ( $LR_{uni}$ ;  $LR_{biv}$ ;  $GF_{uni}$ ;  $GF_{biv}$ ), *Simul* es el procedimiento de simulación (Bootstrap o Monte Carlo) y *Estima* es el método de estimación (POM o NOR)

Para contrastar los efectos incluidos en este modelo se lleva a cabo un análisis de varianza Multivariante (MANOVA). Los resultados incluyen pruebas de significación estadística para cada uno de los efectos y también las de significación práctica  $\eta^2$  (eta cuadrado) que describe la proporción de variabilidad total atribuible a un factor. Los análisis mencionados en este apartado se llevan a cabo con el programa SPSS 12.

### 3.5 Resultados

Se encontró que un menor número de soluciones que converge al utilizar las muestras generadas por Bootstrap en comparación con el procedimiento Monte Carlo con cada método de estimación. De las 200 réplicas para cada método de estimación el Bootstrap falló en 87 (43,5%) bajo el método POM y en 68 (34%) bajo el NOR. Destaca el hecho de que el procedimiento Monte Carlo convergió en todos los casos (99,95%) salvo en uno con el método POM. Las soluciones no convergentes fueron eliminadas y se caracterizaron por la presencia de casos Heywood tanto en los índices  $\chi^2_{LR}$  como en los  $\chi^2_{GF}$ .

La Tabla 3.2 muestra la media y la desviación típica para cada uno de los índices:  $LR_{uni}$ ,  $LR_{biv}$ ,  $GF_{uni}$  y  $GF_{biv}$ . Globalmente los resultados indican que los valores medios de cada índice discrepan sustancialmente de los valores de la solución original con una acentuación más fuerte de estos valores bajo el procedimiento Bootstrap.

El valor de un índice en sí mismo no es suficiente para valorar el ajuste de una solución, por ello, se considera que una diferencia cerrada entre los mismos (univariados, por una parte y bivariados por la otra) es un indicador del buen ajuste. También, se considera que mientras menor sea el valor de los índices mejor será el ajuste de la solución. En consecuencia, en la Tabla 3.2, podemos apreciar cómo los valores de los índices Monte Carlo para el método POM son significativamente más bajos que los suministrados por el Bootstrap, aunque en el

método NOR los valores univariados son más bajos que los valores POM del Monte Carlo, los valores bivariados son sustancialmente mayores.

Tabla 3.2.

Datos simulados: estadísticos descriptivos de índices de ajuste por método de estimación.

Método de Estimación		Bootstrap		Monte Carlo	
		Media	Desv. típ.	Media	Desv. típ.
FIML POM	LR-uni	25,62	49,03	2,69	2,13
	LR-biv	2261,33	488,09	804,04	52,69
	GF-uni	24,51	43,83	2,68	2,13
	GF-biv	2233,41	406,70	753,66	50,19
FIML NOR	LR-uni	4,83	5,25	9,67	8,50
	LR-biv	2187,07	183,84	821,11	71,69
	GF-uni	4,97	5,11	9,81	8,72
	GF-biv	2316,55	209,57	788,87	79,20

La Tabla 3.3 muestra los resultados del ANOVA multivariante que pone a prueba el modelo planteado en la ecuación 5 en todas las replicas y eliminando los casos Heywood. Como se observa, en relación a los coeficientes bivariados ( $LR_{biv}$  y  $GF_{biv}$ ), el modelo explica un 91,6% y 95,2% respectivamente. El efecto del procedimiento de simulación (*simul*) es estadísticamente significativo ( $p = 0,000$ ). El efecto del método de estimación (*estima*) no es estadísticamente significativo ( $p = 0,121$ ) para el índice  $LR_{biv}$ , para el resto de los índices si lo es ( $p = 0,000$ ). La interacción *simul\*estima* no es estadísticamente significativa ( $p = 0,145$ ) para el índice  $GF_{biv}$ , para el resto de los índices sí lo es ( $p = 0,000$ ).

Tabla 3.3

Resultados del ANOVA para los índices  $LR_{uni}$ ,  $LR_{biv}$ ,  $GF_{uni}$  y  $GF_{biv}$

	$LR_{uni}$		$LR_{biv}$		$GF_{uni}$		$GF_{biv}$	
	P	$\eta^2$	P	$\eta^2$	P	$\eta^2$	P	$\eta^2$
Simul	0,000	0,041	0,000	0,902	0,000	0,044	0,000	0,929
Estima	0,000	0,024	0,121	0,004	0,000	0,024	0,000	0,020
Simul*estima	0,000	0,091	0,013	0,010	0,000	0,102	0,145	0,003
Total		0,156		0,916		0,170		0,952

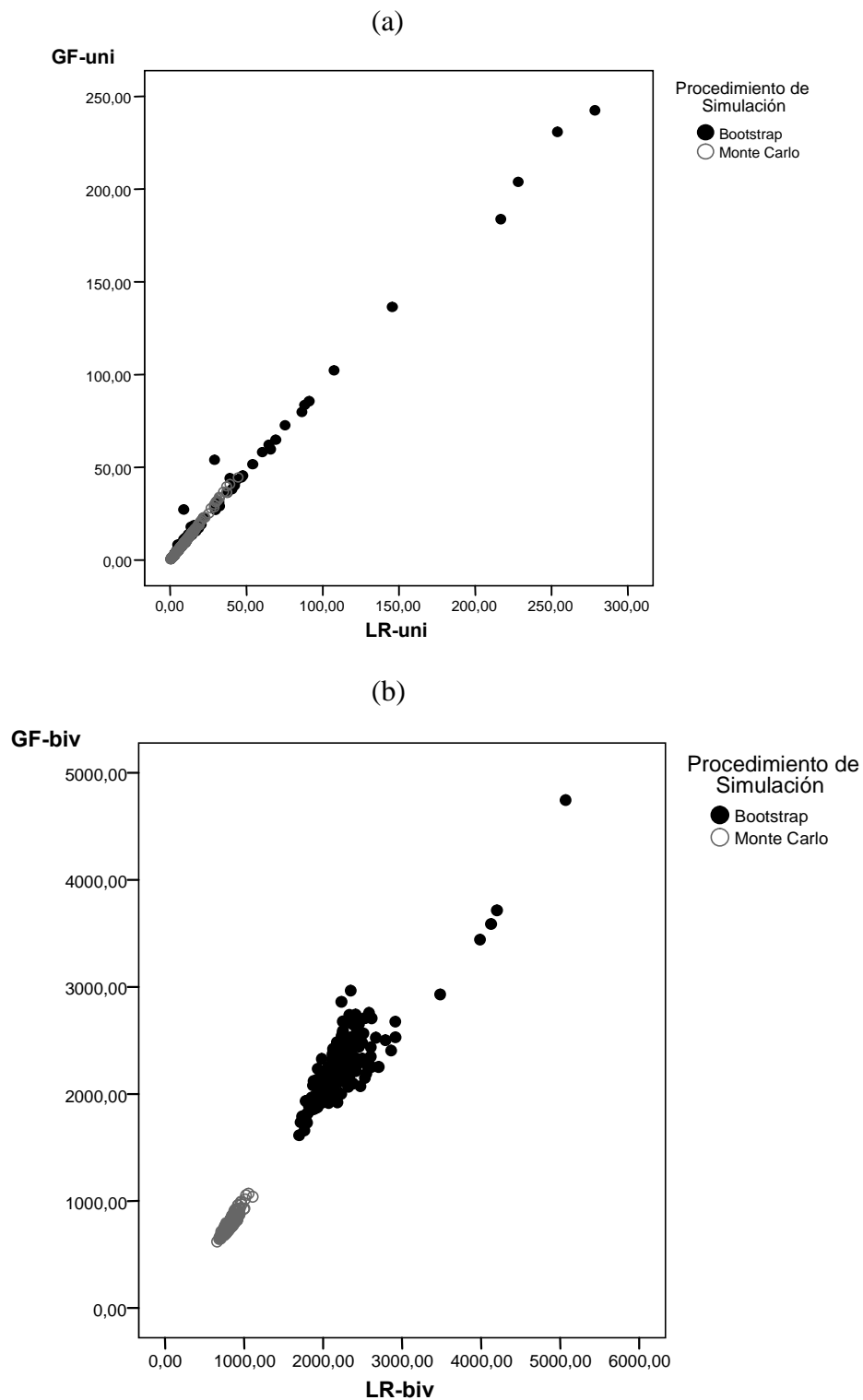


Figura 3.1. Diagramas de dispersión de cada índice por procedimiento de simulación: (a) índices univariados; (c) índices bivariados

El mayor efecto encontrado es el del procedimiento de simulación en todos los índices aunque en menor proporción en los univariados  $LR_{uni}$  y  $GF_{uni}$ . Para los índices bivariados ( $LR_{biv}$  y  $GF_{biv}$ ) el valor de  $\eta^2$  supera el 90%. Con excepción de los efectos de la interacción *simul\*estima* para los índices univariados  $LR_{uni}$  y  $GF_{uni}$  con  $\eta^2$  a cercana a 0,10 el resto de los valores no merecen una atención especial.

Las Figuras 3.1a y 3.1b ponen de manifiesto que el procedimiento Monte Carlo produce estimaciones más bajas y por tanto óptimas en los índices univariados y bivariados. Por otra parte la figura 3.2 también revela que la estimación por la función POM produce valores mayores, esto se debe a los altos valores producidos por la combinación del Bootstrap con el POM, obviando el conjunto de valores para esta combinación los resultados pueden considerarse favorables para la función POM, tanto para los índices univariados (3.2a) como bivariados (3.2b).

En consecuencia, si filtramos los datos por procedimiento de simulación Monte Carlo y método de estimación POM (Figura 3.3), observamos que el rango de los valores para los índices univariados es el más bajo del conjunto y oscila aproximadamente entre 0 y 13 (Figura 3.3a) y para los índices bivariados, el rango de variación va aproximadamente entre 700 y 950. Este conjunto de valores constituyen los mejores resultados de la simulación y permiten establecer, en cada caso, la evaluación del ajuste como buena.

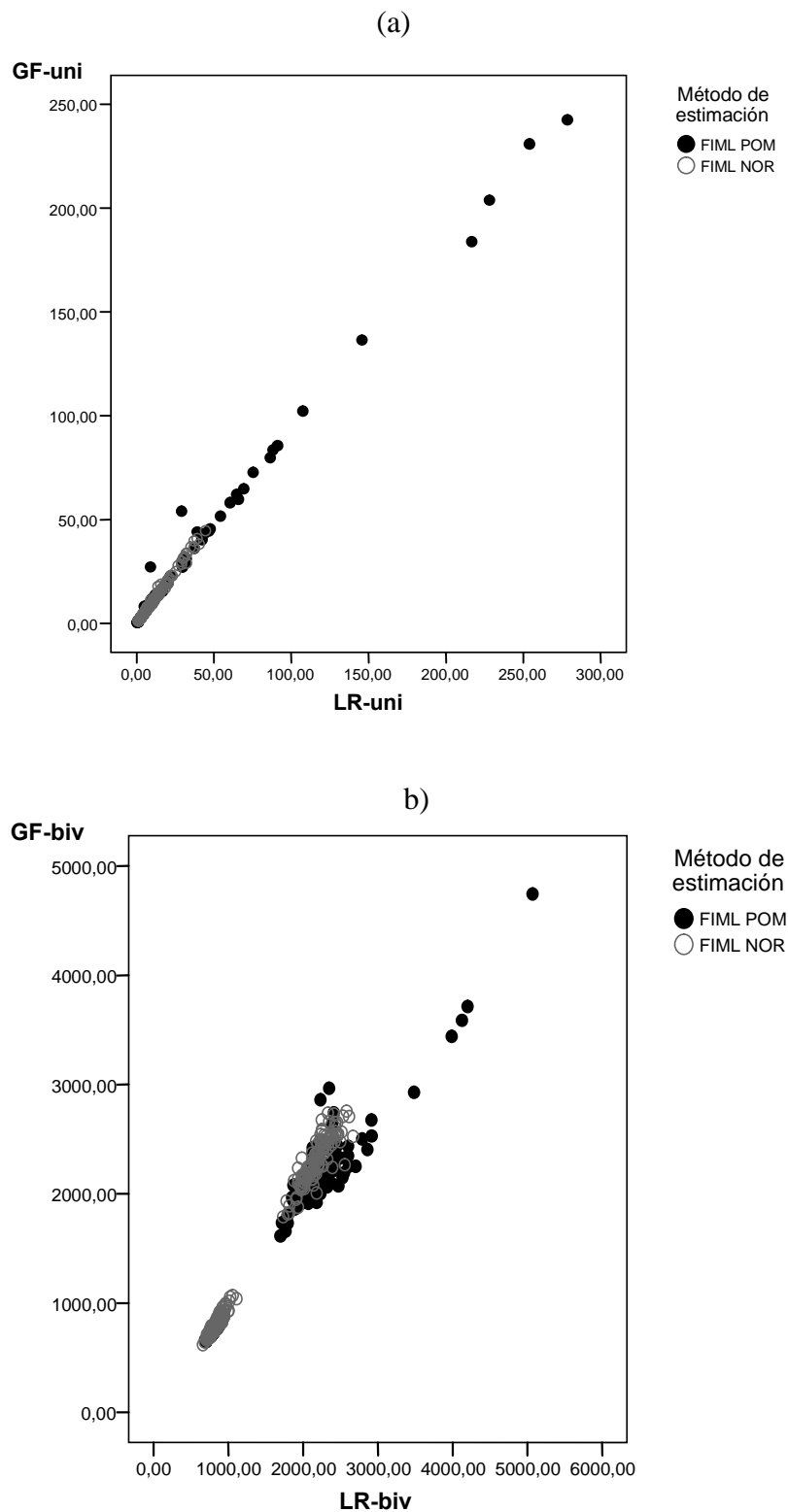


Figura 3.2.- Diagramas de dispersión de índices por método de estimación de estimación: a) Univariados; b) Bivariados

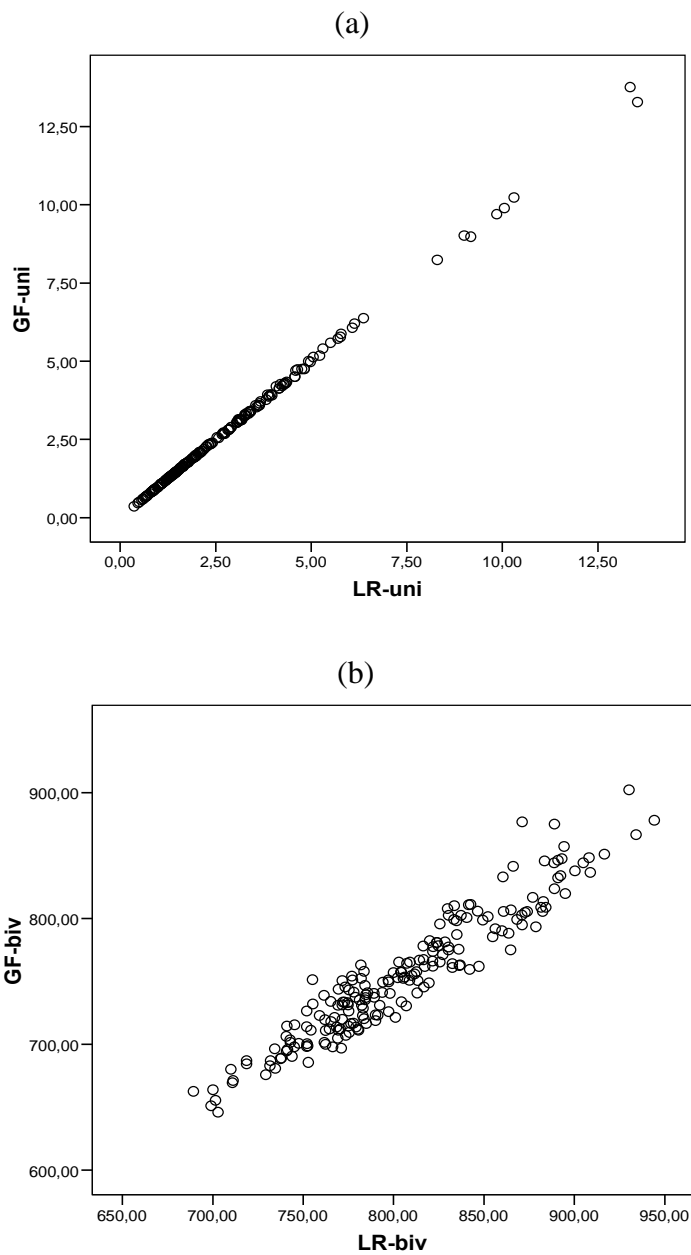


Figura 3.3.- Valores de los índices filtrados por Monte Carlo y POM: a) Univariados; b) Bivariados

### 3.6 Discusión y conclusiones

En este trabajo se sometió a juicio el efecto de las variables: procedimiento de simulación empleado para generar muestras y el método de estimación sobre los índices de ajuste en el contexto de los LVM para variables ordinales bajo la aproximación IRT. Los resultados permiten realizar las siguientes afirmaciones: (1) El procedimiento de simulación empleado para generar las muestras afecta sustancialmente los resultados de los índices de ajuste. (2) Los índices de ajustes en las muestras generadas por Bootstrap coinciden con la presencia de casos Heywood para los estadísticos  $LR_{uni}$  y  $LR_{biv}$  y la falta de convergencia para los estadístico  $GF_{uni}$  y  $GF_{biv}$ , lo que hace necesario que la solución sea eliminada, para el método Monte Carlo la presencia de esta situación es prácticamente nula, estos resultados coinciden con los obtenidos por Fan (1994). (3) En el modelo planteado el factor que tiene una mayor explicación en la variabilidad de los resultados es el procedimiento de simulación. Estos resultados son congruentes con trabajos que hacen alusión a los fallos del remuestreo por Bootstrap: Bollen & Stine(1992) introducen una modificación al *naive bootstrap*; Chernick (1999) enumera una serie de causas por las cuales el Bootstrap falla; Fan (1994) concluye que en el Bootstrap se perpetúan las características de las muestras; Hartmann, (2005) advierte que la simplicidad del Bootstrap puede tentar a los investigadores a usarlo en situaciones inapropiadas; Young (1994) ilustra algunos problemas de las aplicaciones Bootstrap. Otros autores buscan soluciones alternativas Bollen (1990), sugiere usar Bootstrap suministrando como valores iniciales de los parámetros los obtenidos en soluciones previas, lo que de Langeheine et al., (1996), denomina *nonnaive bootstrap* y que se corresponde en esta investigación con el procedimiento empleado por Monte Carlo.

Bajo las condiciones de realización de este trabajo se ha demostrado que la generación de muestras a partir de parámetros previos de investigación y la generación de variables ordinales basadas en la matriz de covarianzas y de los umbrales es mucho más efectiva para la evaluación del buen ajuste que la



obtención de las mismas por remuestreo, por lo cual se sugiere la adopción del procedimiento Monte Carlo, conclusión compartida, aunque en otro contexto con el trabajo de Langeheine et al, (1996).

Finalmente los resultados de este trabajo sugieren que el procedimiento de simulación más efectivo para la generación de muestras de variables ordinales en el contexto de los LTM es el Monte Carlo, pues el Bootstrap falla en un número significativo de veces. En cuanto a la combinación con un método de estimación de información completa en modelos IRT los valores obtenidos bajo POM son significativamente superiores a los obtenidos bajo la función NOR, resultados similares a los encontrados por Jöreskog & Moustaki, (2001). Se trata pues de sugerir el empleo de la simulación de Monte Carlo con la función POM como método de estimación para el tratamiento de datos ordinales en el contexto de variables latentes.

## Capítulo IV

### **Simulación y evaluación del ajuste: aspectos técnicos y computacionales.**

#### **Resumen:**

El programa LISREL 8.8/PRELIS 2.81 permite realizar el Análisis Factorial de datos Ordinales (comando OFA) con métodos de estimación de información completa (FIML) en un conjunto de datos que contiene  $n$  muestras obtenidas por simulación. No obstante, cuando el número de replicas es superior a uno, se produce un error del comando que imposibilita obtener soluciones que se ajusten a funciones normales (NOR) o a funciones logísticas (POM). En este capítulo se propone un procedimiento que permite generar  $n$  muestras y conocer los índices de ajuste para cada replica en modelos de la IRT haciendo posible la ejecución del comando OFA para simulaciones Monte Carlo. Se comparan las soluciones usando la aproximaciones UV (con método de estimación WLS ) y IRT.

## 4.1 Introducción

El software comercial para modelado de ecuaciones estructurales LISREL 8.81, (Jöreskog & Sörbom, 2006a), con su procesador PRELIS 2.81 (Jöreskog & Sörbom, 2006b), tiene disponibles dos técnicas de simulación para extraer muestras aleatorias: el Bootstrap (Efron, 1979, 1982, 1987, 1988; Efron & Tibshirani 1986, 1993) y Monte Carlo (Rubinstein, 1981; Mooney 1997; Skrondal 2000; Paxton, Curran, Bollen, Kirby & Chen, 2001). La diferencia principal que se establece entre ambas técnicas es que en Bootstrap se toman muestras aleatorias de una muestra original de datos reales mientras que en Monte Carlo las muestras aleatorias se pueden generar sin necesidad de disponer de una muestra original de datos (Jöreskog & Sörbom, 2002). Para ambas técnicas, como generador de números aleatorios LISREL y PRELIS utilizan un código en lenguaje FORTRAN de Sharage (1979), el cual genera aleatoriamente distribuciones uniformes con formatos de doble precisión. Los trabajos Fleishman (1979) y de Vale & Maurelli (1983) son utilizados para la generación de datos no normales. Algunos procedimientos Bootstrap para evaluación de parámetros que no están implementados en LISREL, pueden encontrarse en Bollen & Stine (1990; 1992) y diferentes aplicaciones Monte Carlo en LISREL pueden verse, por ejemplo, en Meade & Lautenschlager, 2004; Ximenez & García (2005) y Beaducel & Wittmann (2005).

En el tratamiento de datos ordinales en modelos de variables latentes la última versión del PRELIS 2.81 emplea un comando para el análisis factorial de datos ordinales denominado OFA que utiliza métodos de estimación de información completa (FIML). Las funciones de ojiva normal (NOR) y la logística (POM) son adoptadas como funciones de enlace y ambas son consideradas como buenas aproximaciones a los modelos de la teoría de respuesta al ítem (IRT). El FIML toma los datos brutos como entrada para el cálculo de los parámetros; en consecuencia, los índices de ajuste obtenidos son diferentes a los que se obtienen en modelos de ecuaciones estructurales (SEM) con métodos de estimación de

información limitada (mínimos cuadrados generalizados, GLS y mínimos cuadrados ponderados, WLS), los cuales emplean la matriz de correlaciones policóricas (PM) y la matriz de covarianzas asintóticas (ACM) como entrada de datos. Detalles de esta aproximación teórica, conocida como aproximación de underlying variables (o Aproximación UV) se encuentran en Jöreskog & Moustaki, 2001; Bartholomew, Steele, Moustaki, & Galbraith, 2002; Moustaki (2003). Una descripción detallada y aplicaciones de la aproximación IRT con LISREL se encuentra en Jöreskog & Moustaki (2006).

Los programas LISREL 8.8 y PRELIS 2.81 tienen algunas limitaciones. Por ejemplo, hemos encontrado una falla grave en el software PRELIS al correr un programa en el cual se ejecuta el comando OFA con un conjunto de datos (B1.RAW) para 100 muestras de 600 observaciones cada una usando el comando BS o simulación Monte Carlo:

```
DATA NI=10 NO=600 RP=100  
RA=B1.RAW  
OFA POM NF=2  
OU
```

Esto supone la imposibilidad de ejecutar  $n$  réplicas por simulación para cada muestra utilizando el FIML con las funciones POM y NOR en la aproximación IRT. Esto nos lleva a pensar que el PRELIS está especialmente diseñado para la aproximación UV, pues los índices de ajuste para los modelos IRT no se almacenan en un único archivo como ocurre en la aproximación UV, lo cual no facilita el análisis posterior de sus distribuciones en modelos IRT.

En el presente capítulo se propone un procedimiento de software que usando la simulación Monte Carlo permite el ajuste de LVM bajo la aproximación IRT superando las limitaciones arriba mencionadas. Se trata de un método alternativo a la aproximación UV (Jöreskog & Sörbom, 2002) para la generación de muestras por simulación corriendo el comando OFA con las funciones POM y NOR de la aproximación IRT (limitado a modelos de 1 y 2 parámetros), el cual permite obtener los índices de ajuste y otros parámetros para cada réplica usando PRELIS y otras herramientas comunes de software.

## 4.2 Procedimiento

El procedimiento alternativo a la aproximación UV usando la técnica Monte Carlo es mostrado en la Tabla 4.1. La propuesta es descrita en la segunda columna de la tabla y se basa en la generación de muestras y evaluación de los índices de ajuste para la aproximación IRT. El procedimiento de Jöreskog & Sörbom (2002) implementado LISREL 8.81 and PRELIS 2.81 para la aproximación UV está descrito en la primera columna de la Tabla 4.1.

Las aproximaciones UV e IRT comparten el primer y segundo paso relacionados con la generación de parámetros para ajustar el modelo, que consisten en: 1) Estimar las frecuencias marginales, la matriz de correlaciones policóricas (PM), la matriz de covarianzas asintóticas (ACM) y los umbrales para cada variable ordinal; y 2) Ajustar el modelo usando como datos de entrada la matriz de correlaciones policóricas con el método WLS, obteniendo así la matriz de covarianzas ajustada (CM), las diferencias comienzan al final del segundo cuando usamos el comando OU para generar el formato de entrada de los datos (matrices vs datos brutos), en el tercer paso con el ajuste del modelo, en el cuarto paso con la estimación de parámetros y finalmente con los índices de ajuste calculados.

Tabla 4.1.

Procedimiento Monte Carlo: aproximaciones UV e IRT

Pasos	Aproximación UV: Jöreskog & Sörbom, (2002)	Aproximación IRT : propuesta
1	Estimar las frecuencias marginales, la matriz de correlaciones policóricas (PM), la matriz de covarianzas asintóticas (ACM) y los umbrales de cada variable ordinal.	Estimar las frecuencias marginales, la matriz de correlaciones policóricas (PM), la matriz de covarianzas asintóticas (ACM) y los umbrales de cada variable ordinal.
2	Generar variables ordinales con las distribuciones marginales, los umbrales y la matriz de covarianza ajustada, especificando el tamaño para cada muestral y el número de réplicas. Para cada muestra se generan las matrices PM y ACM.	Generar variables ordinales con las distribuciones marginales, los umbrales y la matriz de covarianza ajustada, especificando el tamaño para cada muestral y el número de réplicas. Para cada muestra se generan datos brutos
3	Analizar el ajuste de las matrices PM y ACM al modelo especificado y para cada muestral generar las medidas de ajuste.	Crear un archive de procesamiento por lotes para correr el commando OFA para ajustar las funciones NOR y POM a cada muestra.
4		Extraer de cada solución ajustada por las funciones NOR y POM las variables que muestren su comportamiento: convergencia de la solución, configuración factorial, la suma de los chi-cuadrado residuales por pares de ítems, los índices de ajuste LR y GF.
5		Analizar la distribución empírica del conjunto de medidas de ajuste.

### 4.3 Implementación:

La implementación de software para el procedimiento Monte Carlo se presenta en la Tabla 4.2. El primer y segundo paso son similares para el método propuesto y para

Tabla 4.2  
Implementación Monte Carlo

Pasos	Aproximación UVS : Jöreskog & Sörbom, (2002)	Aproximación IRT : propuesta.
1	Computing Thresholds and PM and AC Matrices from POLACTION.RAW DA NI=10 MI=6 LA= POLACTION.LAB RA= POLACTION.RAW FO;(6F1.0) OU MA=PM PM= POLACTION.PML AC= POLACTION.ACP TH= POLACTION.TH	Computing Thresholds and PM and AC Matrices from POLACTION.RAW DA NI=10 MI=6 LA= POLACTION.LAB RA= POLACTION.RAW FO;(6F1.0) OU MA=PM PM= POLACTION.PML AC= POLACTION.ACP TH= POLACTION.TH
2	DA NI=10 NO=564 MA=PM LA= POLACTION.LAB PM= POLACTION.PML AC= POLACTION.ACP MO NX=10 NK=2 PH=ST FR LX(1,1) LX(2,1) LX(3,1) LX(4,1) LX(5,1) FR LX(6,1) LX(7,1) LX(10,1) LX(8,2) LX(9,2) OU SI=POLACTION.SIG ME=WLS SD V1-V10 OU IX=2345 MA=PM PM=POLACCTION.PMM AC=POLACTION.ACM XB XT	DA NI=10 NO=564 MA=PM LA= POLACTION.LAB PM= POLACTION.PML AC= POLACTION.ACP MO NX=10 NK=2 PH=ST FR LX(1,1) LX(2,1) LX(3,1) LX(4,1) LX(5,1) LX(6,1) FR LX(7,1) LX(10,1) LX(8,2) LX(9,2) OU SI=POLACTION.SIG ME=WLS SD V1-V10 OU IX=2345 MA=PM ND=0 RA= POLACTION.RAW AC=POLACTION.ACM XB XT
3	Generating 200 Sets of Fit Measures from POLACTION.PMM and POLACTION.ACM DA NI=10 NO=600 MA=PM RP=200 LA= POLACTION.LAB REWIND PM= POLACTION.PMM AC= POLACTION.ACM MO NX=10 NK=2 PH=ST FR LX(1,1) LX(2,1) LX(3,1) LX(4,1) LX(5,1) LX(6,1) FR LX(7,1) LX(10,1) LX(8,2) LX(9,2) OU GF= POLACTION.GFM ME=WLS	1) File syntax SPSS for the separation of samples GET FILE='C:\ REPLICAS\REPLI20095.SAV'. FILTER OFF. use 1 thru 600 /permanent . EXECUTE . SAVE TRANSLATE OUTFILE='C:\REPLICAS\RAW1.dat' /TYPE=TAB /MAP /REPLACE. SAVE OUTFILE='C:\REPLICAS\RAW1.sav' /COMPRESSED. GET FILE='C:\ REPLICAS\REPLI20095.SAV'. FILTER OFF. use 601 thru 1200 /permanent . EXECUTE . SAVE TRANSLATE OUTFILE='C:\REPLICAS\RAW2.dat' /TYPE=TAB /MAP /REPLACE. SAVE OUTFILE='C:\ REPLICAS\RAW2.sav' /COMPRESSED. 2) File batch processing parameter estimation taking into account the PRELIS program for each sample PATH C:\PROGRAM FILES\LISREL 88\ PRELIS28 R1.PR2 R1.OUT -NWIN RENAME BIVFITS.POM BIVFITSR1.POM RENAME MULFITS.POM MULFITSR1.POM PRELIS28 R2.PR2 R2.OUT -NWIN RENAME BIVFITS.POM BIVFITSR2.POM RENAME MULFITS.POM MULFITSR2.POM ..... Programa Prelis R1.PR2 DATA NI=10 MI=6 LABELS FIRMAR BOICOT ASIREU CONPOL ENTNDIN CONMED COLCIU ILEGAL VIOLEN INTERN RAW=RAW1.DAT OFA POM NF=2 OU

Tabla 4.2- Continuación

4	Analyzing 200 Sets of Fit Measures in . POLACTION.GFM DA NI=3 LA;CHI2 P RMSEA RA= POLACTION.GFM FO;(16X,2D13.5/52X,D13.5/////) CO ALL OU MA=KM WP	Extraer de cada solución ajustada con las funciones POM y NOR, las variables que muestren su comportamiento: convergencia de la solución, configuración factorial, las sumas de los chi-cuadrados residuales por pares de ítems, los índices LR y GF.
5	Analyzing 200 Sets of Chi-square Measures in POLACTION.GFM DA NI=1 LA;CHI2 RA= POLACTION.GFM FO;(16X,D13.5/////) RE CHI2 OLD=0-2.83,2.8301-3.82,3.8201- 4.67,4.6701-5.49 NEW=1,2,3,4 RE CHI2 OLD=5.4901-6.35,6.3501-7.28,7.2801- 8.38,8.3801-9.80 NEW=5,6,7,8 RE CHI2 OLD=9.801-12.0,12.001-4711 NEW=9,10 OU	Del paso anterior se obtiene un archivo con los valores de las variables de interés con un número de casos igual al número de réplicas hechas, lo que nos permite analizar la distribución empírica del conjunto de medidas de ajuste .

la aproximación UV, aunque en la salida del segundo paso se generan datos brutos en lugar de matrices con la finalidad de permitir la entrada de datos en el tercer paso con el uso los modelos IRT con métodos de estimación de información completa

En el tercer paso, el uso del comando OFA con replicas mayores que uno es adecuado en la aproximación UV para analizar las matrices obtenidas en el segundo paso. Sin embargo, el comando OFA no corre con replicas mayores que uno en la aproximación IRT con datos brutos. Se propone usar dos herramientas de software para resolver esta situación, la primera permite la separación de los datos generados en el procedimiento Monte Carlo usando el programa PRELIS, que fue almacenada en un solo archivo, y la otra permite ejecutar el comando OFA para cada muestra de datos brutos. Se usa un archivo en lenguaje matrix de SPSS para separar los datos y obtener los conjuntos individuales de datos brutos que pueden ser ejecutados posteriormente con el comando OFA en programa determinado (por ejemplo, R1.PR2) que debe ser ejecutado en un archive de procesamiento por lotes usando el comando RENAME para evitar que las soluciones obtenidas sean sobre escritas.

En el cuarto paso, el comando GF nos permite obtener un archivo con los índices de ajuste de cada solución. Esto sólo es posible en la aproximación UV, mientras que en la aproximación IRT se debe inspeccionar cada solución y extraer



la suma de los chi-cuadrado residuales para pares de ítems y las medidas de ajuste LR y GF univariadas y bivariadas.

#### **4.4 Aplicación:**

##### **4.4.1 Muestra:**

La muestra fue obtenida a partir de la aplicación de una escala de Acción Política a un conjunto de 643 sujetos (47% hombres y 53% mujeres) con edad media de 28 años y desviación típica de 10.2. Los valores perdidos fueron excluidos por “listwise deletion”, es decir, que aquellos patrones de respuesta que tienen un valor perdido en alguno de los ítems son omitidos del análisis obteniendo así una muestra de 564 respondientes. Para la simulación Monte Carlo el tamaño de la muestra fijado fue 600 sujetos.

#### **4.5 Resultados:**

En la Tabla 4.3 vemos que para el WLS el chi cuadrado muestra una amplia dispersión, los índices de ajuste incremental (GFI, AGFI y CFI) muestran valores favorables, sin embargo, el RMSEA se encuentra por encima de 0.05. Para los índices de la aproximación IRT el método POM muestra valores más bajos y por tanto más eficientes en todos los índices, destacándose sustancialmente los índices univariados en comparación con los obtenidos bajo NOR. Las figuras 4.1 y 4.2 muestran la distribución de los índices principales para cada método de estimación. Con respecto a WLS en la figura 4.1a destaca un rango de variación aproximado del chi-cuadrado entre 60 y 180, una desviación estándar alta (17,76) lo que da cuenta de la amplia dispersión de sus valores. Por otra parte el RMSEA (figura 4.1b) muestra una media de 0,065 y número bajo de casos con valores próximos a 0,08.

En la aproximación IRT los resultados son Buenos y similares con ambos métodos de estimación. El método POM proporciona valores entre 650 y 950 con

media de 804.07 para el LR-biv (figura 4.2a) y entre 680 y 900 con media de 753.66 para el GF-biv (figura 4.2b), el método NOR proporciona rangos de variación entre 600 y 1100 para ambos índices (figuras 4.2c y 2d) con medias de 821.11 y 788.87, respectivamente

Table 4.3

Aproximaciones UV y IRT: medias y desviaciones típicas por procedimiento de simulación

Método de Estimación	Índice	Media	DT
WLS	Chi	127.41	17.76
	RMSEA	0.07	0.01
	GFI	0.99	0.00
	AGFI	0.98	0.00
	CFI	0.96	0.01
POM	LR-uni	2.69	2.13
	LR-biv	804.06	52.70
	GF-uni	2.68	2.13
	GF-biv	753.66	50.20
NOR	LR-uni	9.67	8.50
	LR-biv	821.11	71.69
	GF-uni	9.81	8.72
	GF-biv	788.87	79.21

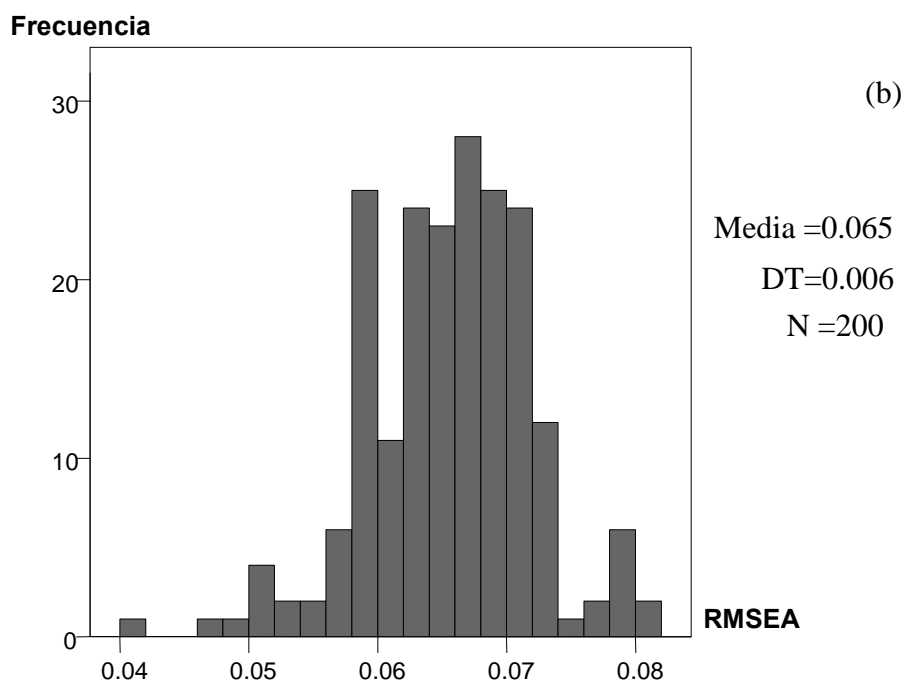
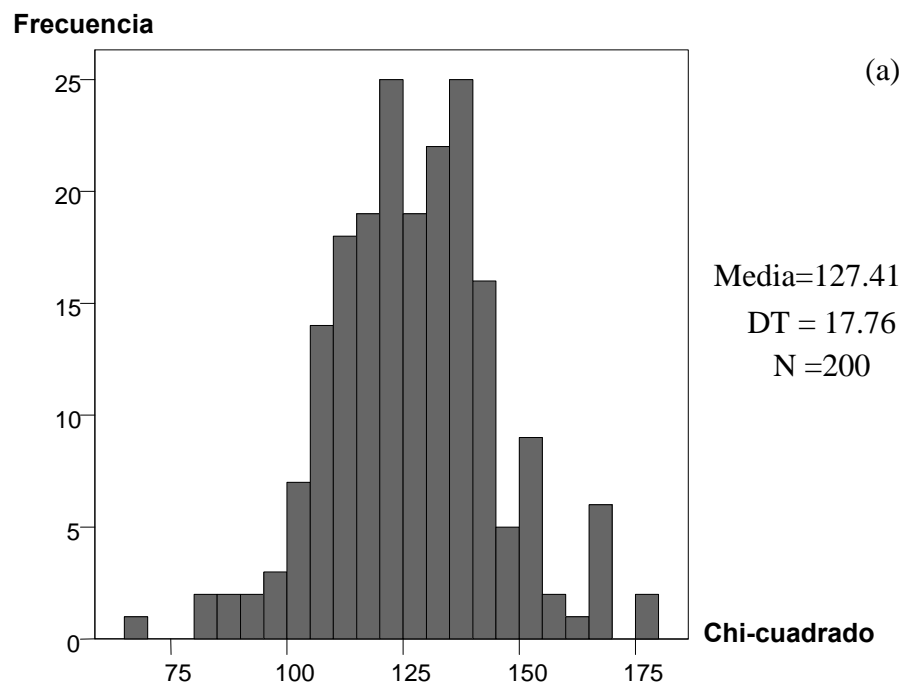


Figura 4.1  
Aproximación UV: resultados de la simulación por método de estimación: a)  $\chi^2$  y  
b) RMSEA para UV-WLS.

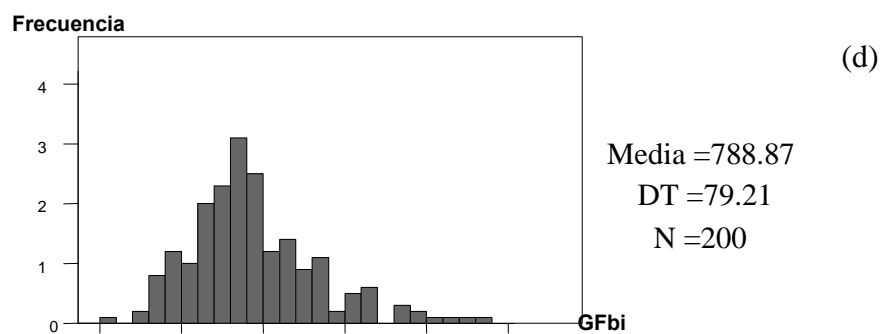
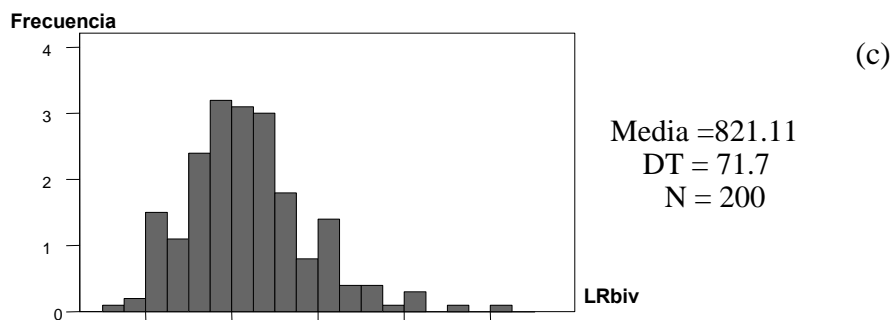
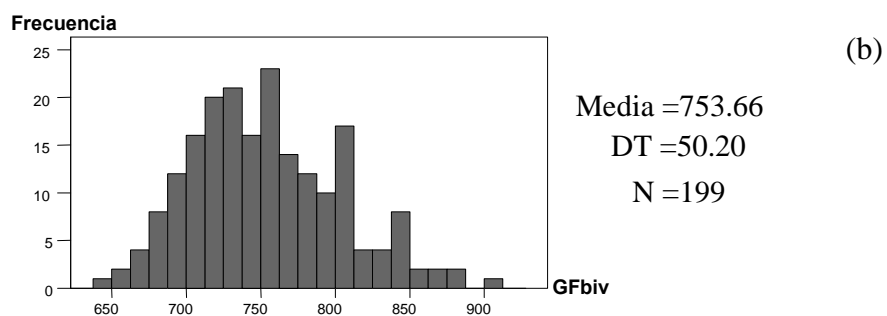
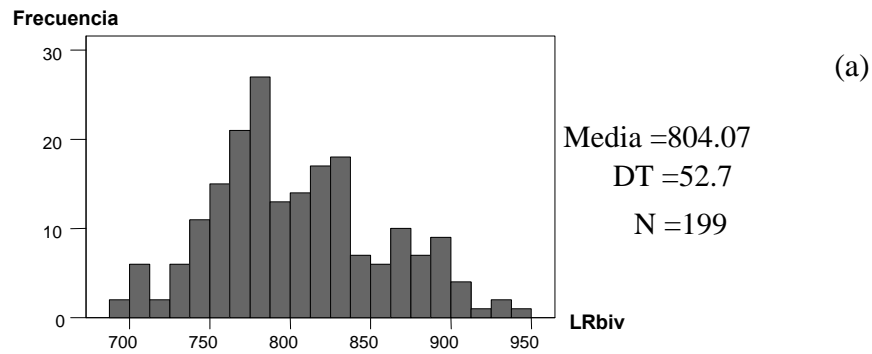


Figura 4.2.

Aproximación IRT: resultados de la simulación por método de estimación:  
a)LRbiv y b) GFbiv para IRT-POM; c)LRbiv y d) GFbiv para IRT-NOR.

**4.6 Conclusiones:**

El análisis de las distribuciones de los índices obtenidas bajo procedimientos de simulación, es común en el análisis de modelos de variables latentes. El software LISREL constituye una herramienta poderosa y ampliamente utilizada por investigadores en esta área. Errores en los códigos son muy comunes en software de diversa índole, como es el caso del comando OFA al intentar realizar más de una réplica.

Las variaciones entre los procedimientos de Jöreskog & Sörbom (2002) y los aquí propuestos se basan en la consideración del tipo de método de estimación: Limited information vs Full Information y, consecuentemente, en el tipo de entrada de datos (matrices vs datos brutos). El procedimiento propuesto permite explotar aun más la información obtenida en las diferentes salidas del programa.

El procedimiento y su implementación descrito en el presente trabajo para Monte Carlo es extensible para el Bootstrap (BS Command) del PRELIS, que está afectado por las mismas limitaciones que se explicaron para la simulación por Montecarlo. En Bootstrap la implementación de los pasos 2, 3, 4 y 5 coinciden con la implementación Monte Carlo con una ligera variación en el paso 1 al generar los conjuntos de datos brutos para cada muestra, esto implica, usar comando OU asignando el número de réplicas (comando BS) y el porcentaje de fracción de muestreo (Comando SF) para definir el tamaño de la muestra. (Ver apéndice A)

Una limitación de este procedimiento, dependiendo del número de réplicas, es que el costo computacional puede ser alto, aunque esto pasa a ser un problema menor con la aparición de ordenadores más potentes en el mercado. Finalmente se pretende que esta alternativa ayude a los investigadores a solventar las limitaciones que hemos mostrado en el programa PRELIS.

## Apéndice A

## Paso 1. Implementation Bootstrap

Pasos	Aproximación UV\$ : Jöreskog & Sörbom, (2002)	Aproximación IRT : propuesta.
1	Generating 200 TM matrices (Kendall's tau-c correlations) by bootstrapping from POLACTION.RAW DA NI=10 MI=6 LA=POLACTION.LAB RA= POLACTION.RAW FO;(6F1.0) OU SP MA=TM BS=200 SF=95 BM=POLACTION.TMB IX=1234567	DA NI=10 MI=6 LABELS FIRMAR BOICOT ASIREU CONPOL ENT DIN CONMED COLCIU ILEGAL VIOLEN INTERN RA=POLACTION.RAW OU SP MA=TM BS=200 SF=95 WI=2 ND=0 RA=BOOT20095.RAW BM=POLACTION.TMB IX=1234567

## Capítulo V

# Un nuevo test para evaluar la bondad de ajuste en tablas de contingencia de amplia dispersión. Comparación con otros índices.

### Resumen:

En el presente capítulo se parte de una exposición del problema de las tablas de contingencia de amplia dispersión con datos ordinales (sparseness). Se examinan los diferentes estadísticos que se usan para evaluar el ajuste en modelos que presentan esta situación. Partiendo de métodos de estimación de estimación completa, y ajustando las funciones POM y NOR desde la IRT, se establece una relación entre las diferentes medidas de ajuste resaltando sus elementos de cálculo, interpretación y ámbitos de aplicación. Se aplica un estudio de simulación para evaluar el comportamiento de estos índices y a partir de los resultados se formulan dos nuevas medidas de ajuste: los coeficientes  $\varphi_{LR}$  y  $\varphi_{GF}$ , que permiten realizar una valoración global del ajuste del modelo, solventando así las carencias en este campo.

### 5.1 Introducción

La bondad del ajuste de modelos de variables latentes con datos ordinales se suele comprobar utilizando índices globales tales como  $\chi^2$  o  $G^2$ , también son considerados para evaluar el ajuste en análisis factorial confirmatorio (Jöreskog, 1969). Ambos estadísticos, bajo la hipótesis nula de no diferencia entre las frecuencias de respuestas observadas y esperadas, siguen aproximadamente una distribución  $\chi^2$  con  $2^p - 2p - 1$  grados de libertad, siendo  $2^p$  el número de patrones (datos binarios) y  $2p$  el número de parámetros a estimar.

La mayor dificultad que presentan los estadísticos basados en  $\chi^2$  que requieren muestras grandes cuando hay un número considerable de ítems con 3 ó 4 categorías de respuestas (Ostini & Nering, 2006). Por ejemplo, con 5 ítems dicotómicos la tabla de contingencia tiene  $2^5$  patrones, por lo que en un test que tuviera 10 ítems, lo cual es un test corto en la práctica, tendría una tabla de  $2^{10}=1024$  patrones. Para aplicar el contraste es necesario que la muestra tenga al menos cinco veces más datos que el número de celdas de la tabla de contingencia por lo que para un test de 10 ítems haría falta una muestra de 5120 sujetos. Esto supera las posibilidades de la mayoría de las aplicaciones reales. Con los modelos politómicos la situación es más difícil, si los ítems tienen  $k$  alternativas la tabla de contingencia tiene  $k^J$  celdas. Por ejemplo para 10 ítems de 4 alternativas, la tabla contiene  $4^{10}=1048576$  celdas y con 5 veces más sujetos que celdas es necesaria una muestra de 5242880, (Abad et al., 2006).

En caso de buen ajuste, las diferencias entre ambos índices ( $\chi^2$  o  $G^2$ ) son muy pequeñas (Moustaki, 2003), pero en caso de sparseness, como se explicó en el párrafo anterior, no es conveniente aplicar ninguno de esos índices (Reiser & Lin, 1999), pues proporcionan índices de ajuste pobres (Kelderman, 1996)

En la teoría de respuesta al ítem (IRT) diversos índices han sido propuestos para evaluar el ajuste en tablas de contingencia de amplia dispersión para datos binarios y ordinales usando métodos de estimación de información completa. Las funciones logística (POM) y normal (NOR) usan como índices de ajuste el  $\chi^2$  likelihood ratio ( $\chi^2_{LR}$ ) y el  $\chi^2$  global fits ( $\chi^2_{GF}$ ), obtenidos a partir de la frecuencia observada y esperada de los patrones de respuestas, para evaluar el ajuste en cada celda o la suma de éstos



( $S_{LR}$ , para el total de los  $\chi^2_{LR}$  and  $S_{GF}$  para el total de los  $\chi^2_{GF}$ ) en tablas de contingencia para cada variable y pares de variables (Drasgow et al., 1995; Chernyshenko et al., 2001; Bartholomew et al., 2002; Bartholomew & Knott, 1999; Jöreskog & Moustaki, 2006; Maydeu-Olivares, 2006).

Un índice alternativo, basado en un familia de parámetros  $\lambda$  cuyos valores son asignados usando información previa, es propuesto por Read & Cressie (1988). Estudios comparativos de simulación muestran las diferencias en los valores de los índices  $\chi^2_{LR}$  y  $\chi^2_{GF}$  y el propuesto por Read & Cressie (1988), Collins et al., (1993). Langeheine et al., (1996) y Collins et al., (1993) plantean el uso de simulación Monte Carlo como un mejor método para la evaluación del ajuste basada en el análisis de las distribuciones de los índices  $\chi^2_{LR}$  y  $\chi^2_{GF}$ . Mavridis et al, 2007 presentan una medida de ajuste para ítems binarios basada en la comparación de los odds ratios y el error tipo I. Otros métodos para evaluar el ajuste en estos modelos son presentados en Bartholomew et al., 2002; Sclove (1987), y están basados en la proporción explicada del ajuste y en criterios para la selección de modelos que incluyen métodos bayesianos.

## 5.2 Problemas resueltos y no resueltos

Diversas investigaciones han permitido implementar reglas prácticas para la interpretación de  $\chi^2_{LR}$  y  $\chi^2_{GF}$  en el ajuste de cada celda en tablas de contingencia univariadas y bivariadas para cada variable y pares de variables (Bartholomew & Tzamourani, 1999; Bartholomew et al., 2002; Jöreskog & Moustaki, 2001; Moustaki, 2000; Bartholomew & Knott, 1999). Estas reglas no están sistematizadas, son imprecisas, resultan arbitrarias y poco funcionales. Se basan en la comparación del valor de los índices  $\chi^2_{LR}$  y  $\chi^2_{GF}$  o  $S_{LR}$  y  $S_{GF}$  con un umbral establecido para considerar el ajuste como adecuado. Esta situación desconcierta a los investigadores a la hora de tomar decisiones acerca de la calidad del ajuste, pues la existencia de algún valor que supere este umbral, es suficiente para cuestionar la solución y considerar el ajuste como inadecuado.

De manera similar, para tablas de contingencia de amplia dispersión con  $m^p$  patrones de respuesta, los índices existentes se pueden aplicar para evaluar la bondad

del ajuste en cada una de las celdas, o para evaluar las distribuciones marginales de las tablas de contingencia para pares de variables, pero no dan información acerca de la evaluación global del modelo. Esto es un gran inconveniente cuando hay que evaluar un número elevado de soluciones.

En LISREL Jöreskog & Sörbom, (2006a) implementan la obtención de la suma de todos los residuales de cada una de las tablas ( $S_{LR}$ , para el total de los  $\chi^2_{LR}$  and  $S_{GF}$  para el total de los  $\chi^2_{GF}$ ) y las resumen en valores únicos,  $LR_{fit}$  para el total de los  $S_{LR}$  y  $GF_{fit}$  para el total de los  $S_{GF}$ , que los autores consideran útiles para la evaluación del ajuste.

De lo anteriormente expuesto, se propone una pirámide de evaluación del ajuste (ver Figura 5.1) que tiene como base la interpretación de la bondad de ajuste en cada celda usando  $\chi^2_{LR}$  o bien  $\chi^2_{GF}$ , en un segundo nivel la evaluación del ajuste en cada tabla de contingencia representada por  $S_{LR}$  and  $S_{GF}$ . En el tercer nivel se encuentran los valores de  $LR_{fit}$  y  $GF_{fit}$  que resumen el comportamiento del ajuste en los niveles anteriores. En el nivel superior se ubican dos nuevos coeficientes globales que permiten valorar el ajuste del modelo,  $\phi_{LR}$  usando  $LR_{fit}$  o bien  $\phi_{GF}$  mediante  $GF_{fit}$ .

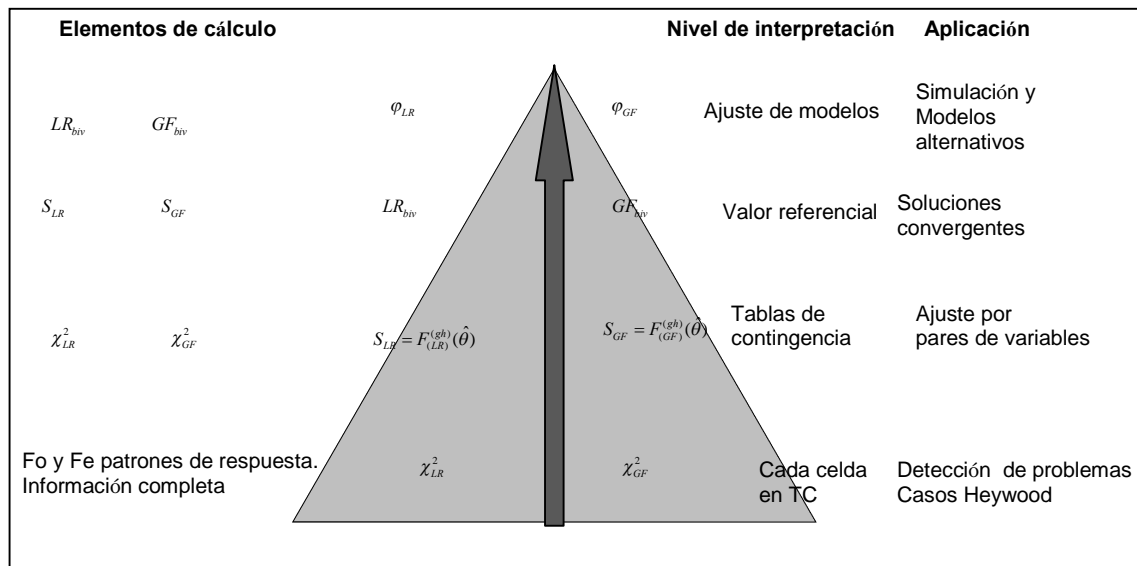


Figura 5.1.

Pirámide de medidas de evaluación del ajuste en tablas de contingencia de amplia dispersión para modelos IRT

La importancia de los coeficientes que se proponen radica en que aprovecha la información de los índices de ajuste existentes para realizar una valoración global del modelo usando un valor único, lo que supera el problema de la inspección individual de cada valor de ajuste por tablas de contingencia y permite agilizar la evaluación del ajuste en un número elevado de soluciones.

### **5.3 Una solución aproximada**

El objetivo del presente capítulo es proponer dos coeficientes globales ( $\phi_{LR}$  y  $\phi_{GF}$ ), calculados a partir de la información suministrada por los índices de ajuste para tablas de contingencia de amplia dispersión con  $m^p$  patrones de respuesta, para evaluar el ajuste en modelos de la IRT con métodos de estimación de información completa (POM y NOR). Asimismo, se establece una regla de interpretación práctica de los nuevos coeficientes basada en un estudio de simulación y a partir de estos resultados se formulan los nuevos coeficientes y se plantean sus reglas de análisis. Finalmente se ejemplifica su aplicación a través de un caso práctico y adicionalmente en el apéndice A se muestra su utilidad aplicándolo con datos utilizados por otros autores (Bartholomew et al., 2002; Jöreskog & Moustaki, 2006).

### **5.4 Método**

Para ilustrar los procedimientos de estimación y ajuste, hemos utilizado los datos obtenidos de la aplicación de una escala de acción política, cuyos ítems y propiedades psicométricas se describen en la sección del instrumento. Los valores estimados son usados como parámetros para el subsiguiente proceso de simulación. Los resultados de la simulación son analizados e interpretados y a partir de éstos se formulan los nuevos estadísticos, se establece una regla práctica para su interpretación y se muestra su aplicación.

#### **5.4.1 Ajuste en tablas de contingencia de amplia dispersión:**

La probabilidad incondicional del patrón de respuestas  $\mathbf{x}_r = (x_1=a_1, x_2=a_2, \dots, x_p=a_p)$  es  $\pi_r(\boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \pi_r(\xi) h(\xi) d\xi$  donde  $h(\xi)$  es la función de densidad de la variable latente  $\xi$ ,  $n_r$  la frecuencia del patrón de respuestas  $\mathbf{x}_r$  y  $p=n_r/N$ , donde  $N$  es el tamaño de la muestra. La estimación de  $\boldsymbol{\theta}$  se hace minimizando la función de ajuste  $F(\boldsymbol{\theta})$  (Joreskog & Moustaki, 2001, 2006):

$$F(\boldsymbol{\theta}) = \sum_r p_r [\ln p_r - \ln \pi_r(\boldsymbol{\theta})] = \sum_r p_r \ln \left[ \frac{p_r}{\pi_r(\boldsymbol{\theta})} \right] \quad (1)$$

Después de minimizar  $F(\boldsymbol{\theta})$  con respecto a  $\boldsymbol{\theta}$  los estadísticos  $\chi^2$  likelihood ratio ( $\chi^2_{LR}$ ) o  $\chi^2$  goodness-of-fit ( $\chi^2_{GF}$ ) pueden ser usados para probar el modelo:

$$\chi^2_{LR} = 2 \sum_r n_r \ln(p_r / \hat{\pi}_r) = 2N \sum_r p_r \ln(p_r / \hat{\pi}_r) = 2NF(\hat{\boldsymbol{\theta}}) \quad (2)$$

$$\chi^2_{GF} = \sum_r \left[ (n_r - N\pi_r)^2 / (N\pi_r) \right] = N \sum_r (p_r - \hat{\pi}_r)^2 / \hat{\pi}_r \quad (3)$$

ambos siguen una distribución  $\chi^2$  con  $d$  grados de libertad:

$$d = (m_1 m_2 - 1) - (m_1 - 1) - (m_2 - 1) - 1 = m_1 m_2 - m_1 - m_2$$

El modelo se evalúa para las distribuciones marginales bivariadas en la muestra, considerando el par de variables  $g$  y  $h$ , siendo  $a$  el número de categorías de la variable  $g$  y  $b$  el número de categorías del variable  $h$ .

Si  $n_{ab}^{(gh)}$  es el número de pares  $(x_g = a, x_h = b)$  y es  $p_{ab}^{(gh)} = n_{ab}^{(gh)} / N$  su frecuencia, la función  $F$ , para el par de variables  $g$  y  $h$  es:

$$S_{LR} = F_{(LR)}^{(gh)}(\hat{\boldsymbol{\theta}}) = 2N \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} p_{ab}^{(gh)} \ln \left[ p_{ab}^{(gh)} / \hat{\pi}_{ab}^{(gh)} \right] \quad (4)$$

se utiliza como medida del ajuste bivariado de las distribuciones marginales de las tablas de contingencia

Alternativamente con el mismo propósito se puede utilizar:

$$S_{GF} = F_{(GF)}^{(gh)}(\hat{\theta}) = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} (p_{ab}^{(gh)} - \hat{\pi}_{ab}^{(gh)})^2 / (\hat{\pi}_{ab}^{(gh)}) \quad (5)$$

#### 5.4.2 Reglas de interpretación:

Las reglas generales para la evaluación del ajuste en tablas de contingencia de amplia dispersión (Jöreskog y Moustaki ,2001; Bartholomew et al., 2002) se pueden resumir como sigue:  $\chi^2_{LR}$  y  $\chi^2_{GF}$  son calculados para cada una de las celdas de las tablas de contingencias y se distribuyen como una  $\chi^2$  con 1 grado de libertad, aquellos valores mayores que 4 indican un pobre ajuste ( $\alpha = 0.05$ ). Para las distribuciones marginales bivariantes de las tablas de contingencia se dan dos reglas diferentes calculando  $S_{LR}$  y  $S_{GF}$ . En un caso, si  $S_{LR}$  y  $S_{GF}$  son valores mayores que  $4m_i m_j$ , el ajuste es pobre, aunque también proponen un límite más conservador es  $3m_i m_j$  ( $m_i$  son las  $m$  categorías de la variable  $i$  y  $m_j$  son las  $j$  categorías de la variable  $j$  en la tabla de contingencia). En el otro caso, se considera que si los valores de  $S_{LR}$  o de  $S_{GF}$  son mayores que el valor de  $\chi^2_{[(m_i m_j)-1]}$  el ajuste es malo ( $\alpha = 0.01$ ).

#### 5.4.3 Sujetos

La muestra estuvo conformada por un total de 643 participantes sujetos (edad media 28 años,  $SD=10.2$ ), el 47% Hombres y 53% Mujeres, cumplieron una escala para la medida de la Acción política. Los valores perdidos fueron excluidos por “listwise deletion”, es decir que aquellos patrones de respuesta que tuvieron un valor perdido en alguno de los ítems, fueron omitidos del análisis, reduciéndose la muestra a 564 participantes.

#### 5.4.4 Instrumento

Para este trabajo usamos una modificación de la escala de acción política, que mide acciones políticas convencionales, del ISSP Citizenship (CIS, 2004; Carton et al.,

2004; Wilson et al., 2005). Las propiedades psicométricas y los ítems de la escala fueron expuestos en el capítulo II.

#### 5.4.5 Software

El ajuste de los modelos se realizó con LISREL 8.8 (Jöreskog & Sörbom, 2006a) y su pre-procesador PRELIS 2.81 (Jöreskog & Sörbom, 2006b), el comando de Análisis Factorial Ordinal (OFA) con las funciones POM y NOR fue empleado para obtener cada una de las soluciones durante el proceso de simulación.

#### 5.5 Simulación

El objetivo del procedimiento de simulación es generar un conjunto de soluciones para evaluar los valores bivariados de  $LR_{fit}$  y  $GF_{fit}$  obtenidos al ajustar modelos de la IRT mediante métodos de estimación de información completa usando las funciones POM y NOR. En el diseño se consideran dos variables independientes, el procedimiento de simulación (1.-Bootstrap; 2.-Monte Carlo) y el método de estimación (1.-POM; 2.-NOR). La variable dependiente son los valores bivariados de  $LR_{fit}$  y  $GF_{fit}$  respectivamente.

Inicialmente el modelo se ajusta a los datos obtenidos con las funciones POM y NOR, a esta solución le denominamos solución empírica. Seguidamente se establecen los procedimientos de simulación para la generación de muestras. En primer lugar se aplica el procedimiento Bootstrap tomando de los datos de la solución empírica una fracción de muestreo del 95%. A partir de la matriz de correlaciones policóricas y de los valores de los umbrales de la solución empírica, con el procedimiento Monte Carlo se generan una muestra de tamaño 600 de 10 variables ordinales no normales usando el PRELIS (preprocesor para LISREL). La generación de números aleatorios está basada en la distribución uniforme implementada en la función URAND del programa la cual implementa un algoritmo en fortran desarrollado por Schrage (1979), este método conocido como generadores de congruencia son los más utilizados para generar valores uniformes (Revueña & Ponsoda, 2003). Las soluciones para cada réplica se obtienen

llevando a cabo un análisis factorial ordinal mediante el comando OFA del PRELIS con métodos de estimación de información completa, ajustando las funciones POM y NOR. En total se han realizado 800 réplicas: los procedimientos de simulación y estimación se repiten 200 veces por cada situación experimental ( $200 \times 2 \times 2$ ). Antes de evaluar los resultados se eliminan las soluciones no-convergentes (según el criterio de LISREL tras 250 iteraciones).

### 5.5.1 Resultados de la simulación:

La revisión individual de cada solución permitió, en primer lugar, verificar la convergencia de la solución, posteriormente, se evalúan los índices  $S_{LR}$  y  $S_{GF}$  y sus totales bivariados  $LR_{fit}$  y  $GF_{fit}$  que se obtienen a partir de las matrices que resumen la información del ajuste por pares de variables. En la Tabla 5.1 se observa que de las 400 réplicas para cada método de estimación el Bootstrap no convergió en 154 de ellas (38.5%), 86 bajo POM (21.5%) y 68 con NOR (17%). El procedimiento Monte Carlo converge en todos los casos salvo en uno bajo el método POM (0.02%). Las soluciones que no convergen se caracterizaron por la presencia de casos Heywood y por valores altos en los totales de índices univariados que conllevan a obtener valores elevados en  $LR_{fit}$  y  $GF_{fit}$  (Figura 5.2a). Las soluciones donde se presenta un mejor ajuste son las que obtienen  $LR_{fit}$  y  $GF_{fit}$  más bajos (Figura 5.2b).

Tabla 5.1.  
Soluciones convergentes método de estimación con Bootstrap

Método de Estimación		Convergencia		Total
		No	Si	
POM	f	86	114	200
	%	43.0%	57.0%	100.0%
NOR	f	68	132	200
	%	34.0%	66.0%	100.0%
Total	f	154	246	400
	%	38,5%	61,5%	100.0%

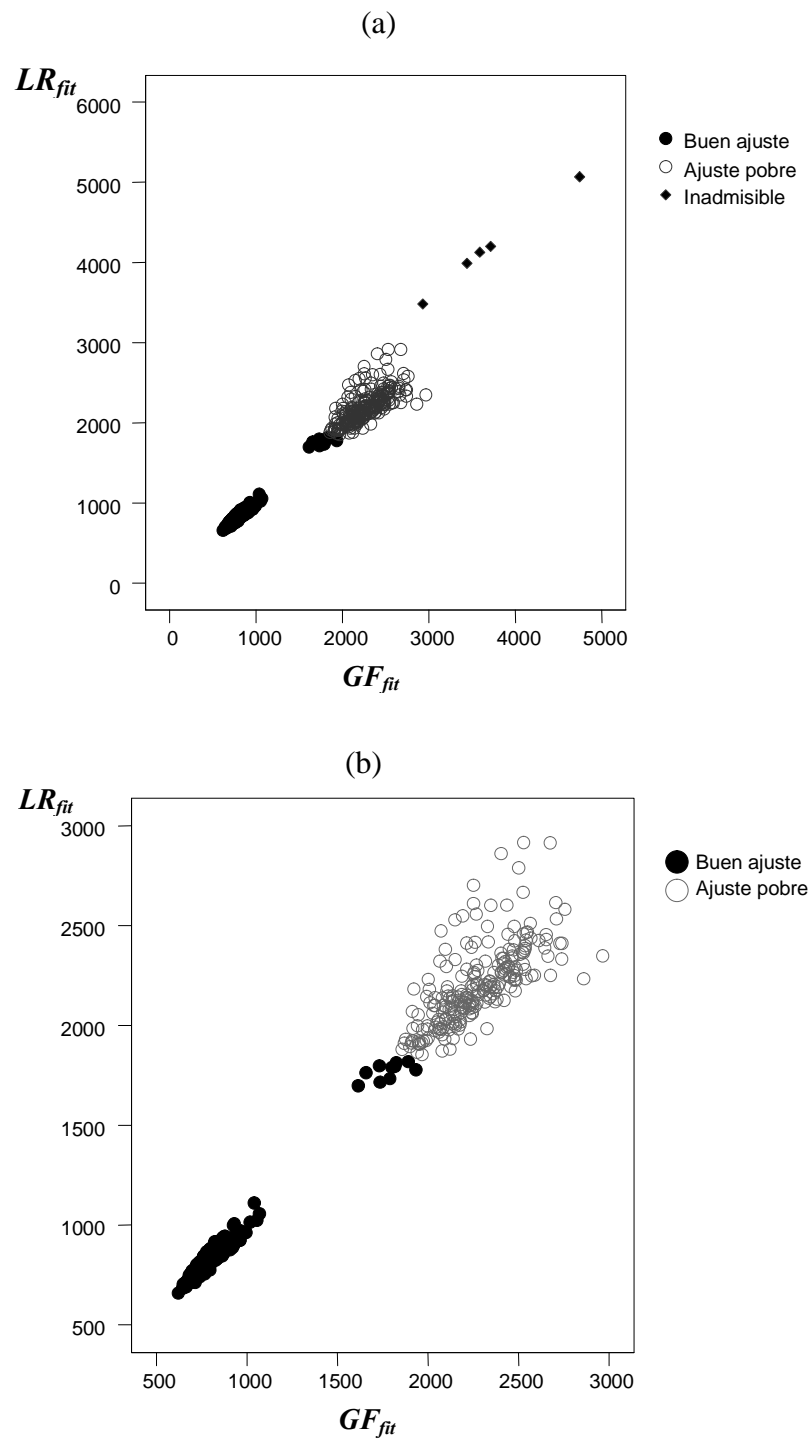


Figura 5.2.

Evaluación del ajuste e índices de ajuste: a) se muestran las soluciones no convergentes con casos Heywood; b) se suprimen los casos Heywood



El número de elementos de la matriz en los que  $S_{LR}$  y  $S_{GF}$  superan el umbral  $\chi^2 m_i m_j$  está correlacionado con los valores de  $LR_{fit}$  y  $GF_{fit}$ , a mayor número de elementos que superan el umbral, mayores son los valores de  $LR_{fit}$  y  $GF_{fit}$ , en consecuencia, la valoración del ajuste (bueno o pobre) está relacionada con el número de elementos de la matriz en los que  $S_{LR}$  y  $S_{GF}$  superan el umbral  $\chi^2 m_i m_j$ .

### 5.6 Formulación de los nuevos coeficientes.

Sea  $S_{LR} = F_{(LR)}^{(gh)}(\hat{\theta})$  la suma de las distribuciones bivariadas por pares de variables, definimos:

$$LR_{fit} = \sum_{i=1}^{\frac{1}{2}n(n-1)} S_{LR} = 2N \sum_{i=1}^{\frac{1}{2}n(n-1)} \left[ \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} p_{ab}^{(gh)} \ln \left( p_{ab}^{(gh)} / \hat{\pi}_{ab}^{(gh)} \right) \right] \quad (9)$$

como la suma de todos los  $n(n-1)/2$   $S_{LR}$  de la matriz que resume la información del ajuste por pares de variables.

Siguiendo las reglas prácticas de Bartholomew, et al. (2002) para considerar un elemento  $S_{LR}$  de la matriz como parte de una solución buena, se tiene que cumplir que:

$$S_{LR} \leq \chi^2 m_i m_j \quad (10)$$

El límite que proponemos para considerar que  $LR_{fit}$  es adecuado viene dado por:

$$LR_{fit} \leq \frac{1}{2} n(n-1) \chi^2 m_i m_j \quad (11)$$

Donde  $n(n-1)/2$  el número de elementos  $S_{LR}$  que contiene la matriz que resume la información para todos los pares de variables,  $\chi^2 m_i m_j$  es el umbral establecido

para aceptar el valor de  $S_{LR}$  como bueno y  $m_i$  son las  $m$  categorías de la variable  $i$  y  $m_j$  son las  $j$  categorías de la variable  $j$  en la tabla de contingencia. Comúnmente  $\chi^2$  toma el valor de 3 ó 4, entonces:

$$\frac{2LR_{fit}}{n(n-1)\chi^2 m_i m_j} \leq \varphi_{LR} \quad (12)$$

El coeficiente de valoración lo definimos como:

$$\varphi_{LR} = \frac{2LR_{fit}}{n(n-1)\chi^2 m_i m_j} \quad (13)$$

Para soluciones convergentes:

$$\varphi_{LR} > 0; \quad \lim_{LR_{fit} \rightarrow \infty} \varphi_{LR} = \infty; \quad \lim_{LR_{fit} \rightarrow \frac{1}{2}n(n-1)\chi^2 m_i m_j} \varphi_{LR} = 1; \quad \lim_{LR_{biv} \rightarrow 0} \varphi_{LR} = 0.$$

De forma análoga, para  $S_{GF} = F_{(GF)}^{(gh)}(\hat{\theta})$

$$GF_{fit} = \sum_{i=1}^{\frac{1}{2}n(n-1)} S_{GF} = \sum_{l=1}^{\frac{1}{2}n(n-1)} \left[ \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} (p_{ab}^{(gh)} - \hat{\pi}_{ab}^{(gh)})^2 / (\hat{\pi}_{ab}^{(gh)}) \right] \quad (14)$$

y el coeficiente que se propone es:

$$\varphi_{GF} = \frac{2GF_{fit}}{n(n-1)\chi^2 m_i m_j} \quad (15)$$

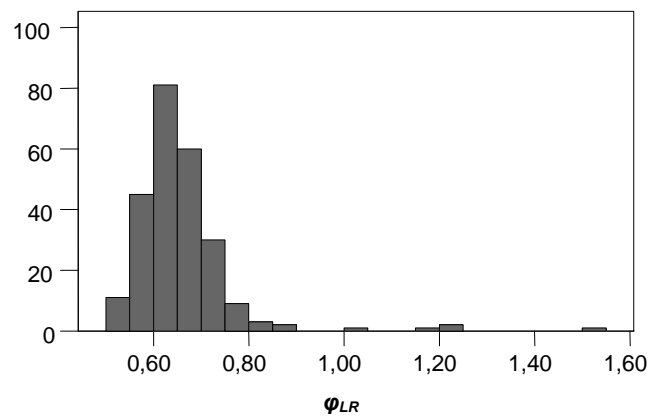
## 5.7 Aplicación

A partir de los resultados obtenidos durante el proceso de simulación se calculan los coeficientes  $\varphi_{LR}$  y  $\varphi_{GF}$ , se analizan sus valores y se propone una regla para su interpretación.

Las Figuras 5.3a y 5.3b muestran la distribución de los índices  $\varphi_{LR}$  y  $\varphi_{GF}$ , estableciendo como umbral  $3m_i m_j$  con el procedimiento Bootstrap. El rango de variación de estos coeficientes va desde 0.50 hasta 1.50 para  $\varphi_{LR}$  y de 0.48 hasta 1.41 para el  $\varphi_{GF}$ . Al inspeccionar individualmente cada solución, aquellas con índices entre 0.50 y 0.53, ambos inclusive, se observó que su ajuste era bueno y ningún elemento de la matriz superaba el umbral  $3m_i m_j$ . En el intervalo comprendido entre 0.54 y 0.57 se encontró que las soluciones incluían al menos 1 elemento que superaba el umbral, sin embargo estos eran muy pocos (entre 1-5) y muchos de estos valores están próximos al umbral considerado. Examinando detalladamente cada valor de las soluciones comprendidas en este rango, se encontró que la correlación entre un par específico de variables, o en concreto la correlación de una variable dada con el resto, afecta el valor del ajuste. A partir de 0.58 se encontró que el número de elementos que supera el umbral se incrementa de forma proporcional al valor del índice. A partir de 0.58 se encontró que el número de elementos que supera el umbral se incrementa de forma proporcional al valor del índice

(a)

Frecuencia



(b)

Frecuencia

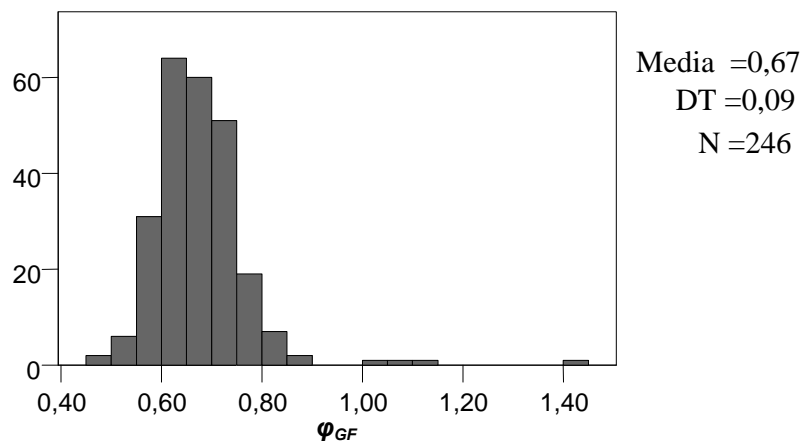
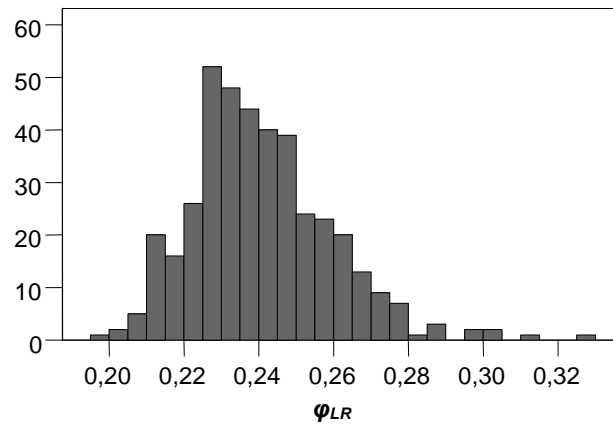


Figura 5.3

Bootstrap: distribución de los coeficientes  $\varphi_{LR}$  y  $\varphi_{GF}$  a)  $\varphi_{LR}$ ; b)  $\varphi_{GF}$

(a)

Frecuencia



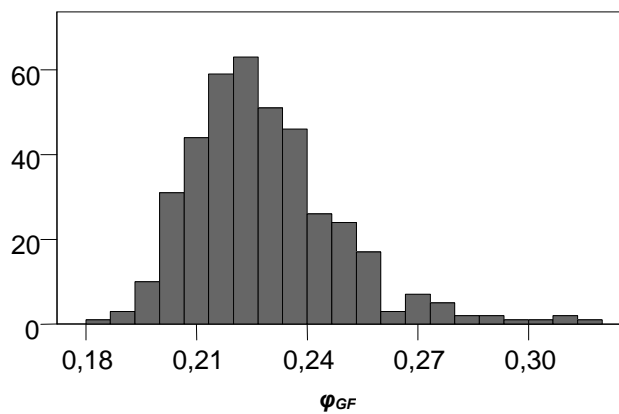
Media =0,24

DT=0,02

N =399

(b)

Frecuencia



Media =0,23

DT = 0,02

N =399

o

Figura 5.4

Monte Carlo: distribución de los coeficientes de ajuste: a)  $\varphi_{LR}$ ; b)  $\varphi_{GF}$ 

Las figuras 5.4a y 5.4b muestran los valores de los índices  $\varphi_{LR}$  y  $\varphi_{GF}$  con el procedimiento Monte Carlo. El rango de variación de los índices va desde 0.20 a 0.33 para  $\varphi_{LR}$  y de 0.18 a 0.32 para  $\varphi_{GF}$ . Al inspeccionar individualmente cada solución se

encontró que su ajuste era óptimo y que la diferencia entre los valores entre los coeficientes  $\varphi_{LR}$  y  $\varphi_{GF}$  era muy pequeña al igual que la diferencia entre  $LR_{fit}$  y  $GF_{fit}$ .

La Figura 5.5 muestra el diagrama de dispersión para los índices  $\varphi_{LR}$  y  $\varphi_{GF}$  por procedimiento de simulación, en ella puede observarse que los valores de obtenidos por Monte Carlo se manejan aproximadamente entre 0.20 y 0.35 lo que indica que las soluciones obtenidas bajo este procedimiento tienen valores de ajuste más próximos a cero y en el caso de Bootstrap se observa un cúmulo de observaciones alrededor del valor 0.60 lo que indican valores de ajuste más distanciados de cero, mostrándose incluso valores en torno a 1 y superiores a éste.

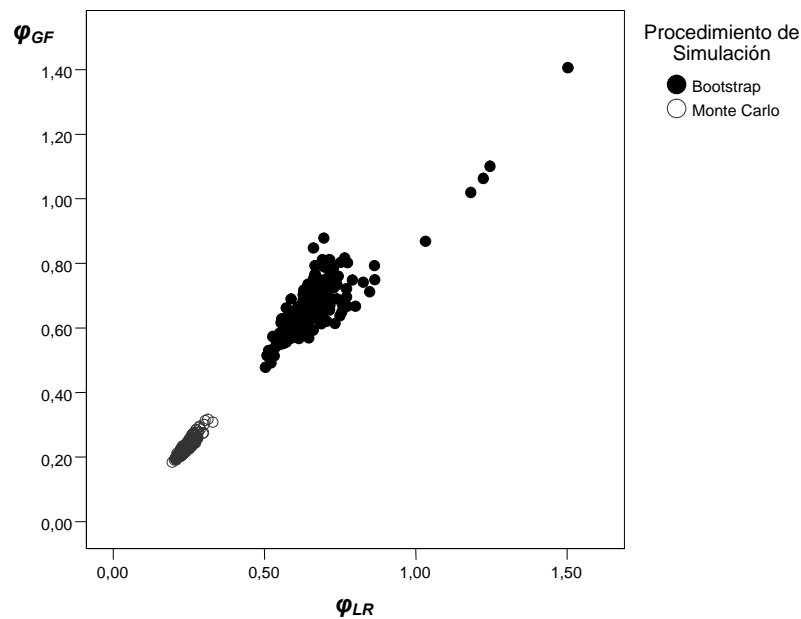


Figura 5.5

Valores del coeficiente  $\varphi_{LR}$  y  $\varphi_{GF}$  con  $\chi^2 m_i m_j = 3 \times 5 \times 5$  para cada procedimiento de simulación

La Figura 5.6 muestra el diagrama de dispersión para los índices  $\varphi_{LR}$  y  $\varphi_{GF}$  por método de estimación, en general se observan valores más próximos a cero con el método NOR y valores superiores a 1 en el método POM, esto se explica por las

soluciones con un pobre ajuste que contienen valores superiores a 1 obtenidas por Bootstrap bajo este método.

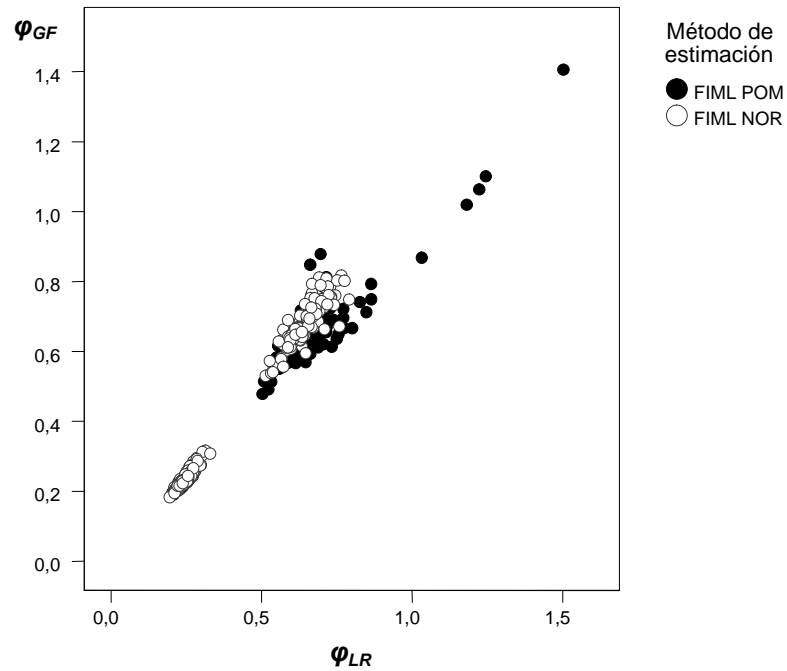


Figura 5.6

Valores del coeficiente  $\varphi_{LR}$  y  $\varphi_{GF}$  con  $\chi^2 m_i m_j = 3 \times 5 \times 5$  para cada método de estimación

Estos resultados nos permiten formular una regla aproximada para la interpretación de  $\varphi_{LR}$  y  $\varphi_{GF}$  en la que en primer lugar valores de  $\varphi \geq 1$  se consideran soluciones inadmisibles. En el intervalo  $0 \leq \varphi \leq 1$  se establecen las siguientes categorías de interpretación: (1) si  $0 \leq \varphi < 0.34$  el ajuste es óptimo; (2) si  $0.35 \leq \varphi < 0.44$  la solución es excelente; (3) si  $0.45 \leq \varphi < 0.53$  la solución es buena; (4) si  $0.54 \leq \varphi < 0.57$  la solución debe revisarse y considerar la exclusión de una variable o un par de éstas para mejorar el ajuste; (5) si  $0.58 \leq \varphi < 0.99$  el ajuste es pobre.

Los resultados la regla de interpretación propuesta son mostrados en la Tabla 5.2. Los valores en ambos índices por categorías de la regla de interpretación son muy similares y apenas presentan ligeras variaciones en algunas de ellas que son justificadas por las pequeñas diferencias entre  $LR_{fit}$  y  $GF_{fit}$ . El porcentaje más alto con 62% está representado por la categoría óptimo lo cual tiene una clara explicación influenciada por

el procedimiento de simulación seleccionado. Dentro del rango de soluciones excelentes no cae ninguna solución. Un número bajo de soluciones (6 casos) caen en la categoría de ajuste bueno. Aproximadamente un 34% está en el rango de soluciones pobres y sólo 5 casos con soluciones superiores a 1 consideradas como inadmisibles para el  $\phi_{LR}$  y 4 para el  $\phi_{GF}$ . En el apéndice A su muestra la aplicación de  $\phi_{LR}$  y  $\phi_{GF}$  para datos publicados.

Tabla 5.2

Valores del coeficiente  $\phi_{LR}$  y  $\phi_{GF}$  con  $\chi^2 m_i m_j = 3*5*5$  según la regla de interpretación

	$\phi_{LR}$		$\phi_{GF}$	
	f	%	f	%
Inadmisible	5	0.8	4	0.6
Optimo	399	61.9	399	61.9
Bueno	6	0.9	6	0.9
Revisar	20	3.1	19	2.9
Pobre	215	33.3%	217	33.6
Total	645	100.0%	645	100.0

## 5.8 Conclusiones

Los coeficientes  $\phi_{LR}$  y  $\phi_{GF}$  suponen un nuevo aporte en cuanto al cálculo e interpretación de los niveles de ajuste en tablas de contingencia de amplia dispersión, pues se obtiene mediante una secuencia de cálculo de los coeficientes ya conocidos que se basan en el análisis del ajuste en celdas y tablas de contingencia por pares de variables. Podemos decir que se trata de una evolución en los estadísticos de ajuste y de una mejora manifiesta en la interpretación de los mismos que se traduce en un aspecto importante a la hora de considerar el uso de LTM con ordinal data.

Cómo se ha mostrado los coeficientes  $\phi_{LR}$  y  $\phi_{GF}$  permiten la valoración global de soluciones con datos ordinales en tablas de contingencia de amplia dispersión con  $m^p$  patrones de respuesta a partir de un valor único, la facilidad de su implementación, cálculo e interpretación resultan útiles en el ajuste de modelos IRT con las funciones POM y NOR.

Los coeficientes  $\phi_{LR}$  y  $\phi_{GF}$  son acotados inferiormente cuando toman el valor de cero, cuanto más próximo se encuentre el valor calculado de  $\phi_{LR}$  y  $\phi_{GF}$  a esa cota



inferior mejor será el ajuste de la solución. Por otra parte su máximo es inalcanzable, pues  $\lim_{LR_{biv} \rightarrow \infty} \varphi_{LR} = \infty$ , por lo tanto su distribución es desconocida y se hace necesaria la implementación de una regla de interpretación estableciendo el valor uno como cota superior para el análisis de los mismos.

Una ventaja que presentan los nuevos coeficientes es que se plantea un rango en que las soluciones pueden ser revisadas para mejorar el ajuste a partir de la inspección del comportamiento de una o más variables o de casos concretos (por ejemplo: casos extremos o Heywood) cuya presencia afectan sustancialmente el ajuste del modelo.

Los ejemplos de aplicación de los nuevos coeficientes mostrados en el apéndice A evidencian la sencillez de su cálculo y la congruencia con las conclusiones sobre la evaluación del ajuste presentada por otros autores. En investigaciones futuras sería interesante evaluar la aplicabilidad de estos índices en el contexto de modelos de clases latentes para ítems binarios y politómicos así como la realización de estudios basados en la transformación de éstos coeficientes para el análisis de sus distribuciones.

## APENDICE A

Datos publicados en Bartholomew et al., 2002, pp.228, 786 respondientes en 1996 a la British Social Attitudes Survey.

(a) Suma de los chi-cuadrado residuales por pares de ítems

	Item 1	Item 2	Item 3	Item 4
Item 1	40,24	22,87	7,87	14,42
Item 2		26,27	16,58	24,48
Item 3			9,66	17,96
Item 4				15,75

$$LR_{biv} = 251,37$$

$$\chi^2 m_i m_j = 3 * 4 * 4 = 48$$

$$n(n-1)/2 = (5 * 4) / 2 = 10$$

$$\phi_{LR} = \frac{2LR_{biv}}{n(n-1)\chi^2 m_i m_j} = \frac{2 * 251,37}{5 * 4 * 48} = \frac{502,48}{960} = 0,523$$

Aplicando la regla de interpretación propuesta, el ajuste es bueno lo que se contrasta que la inspección individual de cada elemento de la matriz que no supera el umbral 48 y coincide con la interpretación dada por el autor a la solución

(b) Bartholomew et al., 2002, ejemplo de la página 226:

	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6
Ítem 1	13,46	4,95	7,92	2,51	1,69
Ítem 2		5,02	8,38	3,76	4,05
Ítem 3			4,47	8,17	12,96
Ítem 4				2,06	6,03
Ítem 5					6,08

$$LR_{biv} = 94,17$$

$$\chi^2 m_i m_j = 3 * 3 * 3 = 27$$

$$n(n-1)/2 = (6 * 5) / 2 = 15$$

$$\phi_{LR} = \frac{2LR_{biv}}{n(n-1)\chi^2 m_i m_j} = \frac{2 * 94,17}{6 * 5 * 27} = \frac{188,34}{810} = 0,2325$$

Aplicando la regla de interpretación el ajuste es óptimo, lo que se contrasta que la inspección individual de cada elemento de la matriz que no superan el umbral 27, de igual manera coincide con la valoración hecha por los autores.

(c) Datos sobre eficacia política: Jöreskog & Moustaki (2006, p. 15)

	NOSAY	VOTING	COMPLEX	NOCARE	TOUCH	INTEREST
NOSAY	0,413					
VOTING	184,764	0,1				
COMPLEX	64,099	37,784	0,452			
NOCARE	54,19	73,144	51,35	1,541		
TOUCH	71,781	51,372	76,72	61,496	0,674	
INTEREST	65,595	68,758	56,576	83,475	60,356	1,006

$$LR_{uni} = 4,186$$

$$LR_{biv} = 1061,46$$

$$\chi^2 m_i m_j = 3 * 4 * 4 = 27$$

$$n(n-1)/2 = (6 * 5) / 2 = 15$$

$$\phi_{LR} = \frac{2LR_{biv}}{n(n-1)\chi^2 m_i m_j} = \frac{2 * 1061,46}{6 * 5 * 48} = \frac{2122,92}{1440} = 1,47$$

	NOSAY	VOTING	COMPLEX	NOCARE	TOUCH	INTEREST
NOSAY	0,412					
VOTING	248,05	0,1				
COMPLEX	92,201	40,93	0,447			
NOCARE	89,051	95,193	53,171	1,52		
TOUCH	119,897	62,14	99,83	133,086	0,651	
INTEREST	94,25	85,354	63,741	254,082	81,219	0,971

$$LR_{uni} = 4,100$$

$$LR_{biv} = 1612,195$$

$$\chi^2 m_i m_j = 3 * 4 * 4 = 27$$

$$n(n-1)/2 = (6*5)/2 = 15$$

$$\phi_{GF} = \frac{2GF_{biv}}{n(n-1)\chi^2 m_i m_j} = \frac{2*1612,19}{6*5*27} = \frac{3224,38}{810} = 3,98$$

Los índices revelan soluciones inadmisibles con un ajuste considera muy malo o pobre por los autores. Los resultados están fueras de los límites de interpretación de los coeficientes  $\phi_{LR}$  y  $\phi_{GF}$ .

## Conclusiones

Los diferentes tópicos que se abordaron y analizaron sobre los LVM y LTM permiten establecer un conjunto de ideas que resultan adecuadas para el tratamiento de datos ordinales en estos modelos. En primer lugar, se ha demostrado que no resulta indiferente la adopción de una u otra aproximación teórica (UV o IRT) para ajustar modelos con datos ordinales. Los resultados de las aplicaciones realizadas muestran un mejor comportamiento en la estimación y evaluación del ajuste de los LTM ajustados desde la IRT.

Por otra parte, se aclaró que el método de estimación empleado afecta significativamente la evaluación del ajuste, debiendo distinguirse entre métodos de información completa y limitada. Se determinó, para la escala empleada, que ajustando las funciones normal (NOR) y logística (POM) desde la IRT se encuentran mejores resultados en la evaluación del ajuste que empleando el GLS y WLS desde la aproximación UV. Bajo condiciones simuladas, se resalta la estabilidad de los métodos ajustados desde la IRT e inconsistencias bajo la aproximación UV. Tomando en cuenta que estos resultados se derivaron a partir de la evaluación del ajuste, son similares a los obtenidos por otros autores en

trabajos en los que también se consideran la estabilidad de los parámetros, (Moustaki et al., 2004; Maydeu-Olivares., 2005).

Al existir diversas alternativas, teóricas y metodológicas, para el tratamiento de los datos ordinales, los investigadores deberían evaluar sus datos y verificar sus distribuciones. La presencia de no normalidad puede ser abordada aproximando los datos con funciones logísticas desde la IRT, pues como se demostró en esta investigación, estos modelos ofrecen mejores resultados en comparación con los datos con métodos lineales.

Entre el conjunto de factores que afectan la bondad del ajuste, se ha señalado que el procedimiento de simulación empleado en los procesos de validación de los índices, tiene un peso significativo en la valoración de los mismos. Al evaluar el efecto conjunto del procedimiento de simulación y los métodos de estimación en LTM tratados desde la IRT, se demostró que el empleo de técnicas de remuestreo, específicamente la implementación no paramétrica de Bootstrap (tal y como está disponible en el software comercial), no resulta adecuada para la evaluación de los índices de ajuste en estos modelos. Lo contrario sucede con las implementaciones Monte Carlo, basadas en la generación de datos ordinales a partir de las matrices policóricas y de covarianzas asintóticas y de los valores de los umbrales para cada ítem, a través del cual se obtienen soluciones convergentes y óptimas que permiten evaluar la precisión de los índices de ajuste. Aunque estos resultados hacen referencia directamente a los efectos sobre la bondad del ajuste, están en línea con los obtenidos por Hartmann (2005); Langeheine et al., (1996) y Young (1994), sobre la estimación de parámetros.

En cuanto al método de estimación de información completa (adoptando las funciones NOR y POM), los pesos de los factores en el modelo evaluado no resultaron con diferencias significativas, lo que implica un comportamiento similar en los métodos de estimación aproximados desde la IRT. Estos resultados son congruentes con los obtenidos en el capítulo II en el que se compararon

métodos de información limitada y de información completa, y en donde a favor de los últimos, se concluía acerca de los óptimos ajustes obtenidos. Por otra parte se sugiere que, en el tratamiento de datos ordinales, se considere la aplicación de métodos de estimación que aproximen las funciones logística y normal desde la IRT, tal y como lo plantean otros autores (Jöreskog & Moustaki, 2001; 2006).

Existen muchas limitaciones en cuanto a la disponibilidad de herramientas de software que permitan obtener índices de ajuste adaptados al comportamiento de datos ordinales en LTM, tratados desde la IRT. Por lo que se opta por la creación de nuevos códigos de programas y la utilización simultánea de herramientas. El procedimiento alternativo propuesto para la generación de muestras por simulación, empleando tanto Bootstrap como Monte Carlo, permite obtener los índices de ajuste para un conjunto de muestras y evaluar la distribución de los mismos, utilizando herramientas de software alternativas al LISREL/PRELIS (Jöreskog & Sörbom, 2006), que permiten superar las limitaciones en este campo. El procedimiento se caracteriza por su funcionabilidad y portabilidad, permitiendo, a través de la implementación de un conjunto detallado de instrucciones, obtener resultados que permitan evaluar el ajuste.

El tema de los índices de bondad del ajuste en LVM representa un aporte especial de este trabajo, pues a partir de la polémica planteada sobre el tema, que versa principalmente sobre lo inadecuadas que resultan las medidas generales de evaluación del ajuste  $\chi^2$  o  $G^2$  para evaluar escalas que presentan un número elevado de ítems categóricos y que cuestiona que estas medidas no dan información acerca de ítems anómalos, se analiza su utilidad y aplicabilidad para la evaluación de modelos específicos y posteriormente se proponen nuevas medidas que superan estas limitaciones.

Considerando estas limitaciones, en el presente trabajo se examinaron las distintas medidas alternativas de ajuste que permitían valorar los modelos, indagando acerca de sus elementos de cálculo y condiciones de aplicación. Este

análisis permitió la formulación de nuevas medidas de ajuste ( $\phi_{LR}$  y  $\phi_{GF}$ ) cuya aplicación proporciona una serie de beneficios: (1) Obtener un valor único que permite evaluar con precisión el ajuste global del modelo; (2) Los valores obtenidos son interpretables a partir de una regla práctica que permite establecer un límite entre buenos y pobres ajustes. Esta información se traduce en una herramienta valiosa para el investigador pues puede valorar la viabilidad de la solución obtenida; (3) Existe un rango de interpretación que sugiere la presencia de ítems anómalos. Esto permite que al investigar, se pueda evaluar el comportamiento de un ítem (o par de ellos) verificando la presencia de casos Heywood y, considerar si con la eliminación del mismo se encontraría un mejor ajuste; (4) Al no ser necesaria la inspección individual del ajuste por ítem (pares o tríos), los nuevos coeficientes resultan apropiados cuando se necesita evaluar un número elevado de soluciones, como las obtenidas en un estudio de simulación o en la evaluación de modelos alternativos. Por lo tanto, el análisis de sus distribuciones facilita las conclusiones acerca de la evaluación del ajuste del modelo; (5) Su cálculo es simple, lo que facilita su implementación mediante software.

En el presente trabajo se hace uso del programa PRELIS para obtener los valores de  $S_{LR}$ ,  $S_{GF}$ ,  $LR_{biv}$  y  $GF_{biv}$ , que son necesarios para su cálculo, pero resulta viable el desarrollo de un programa independiente o la implementación de pequeñas rutinas en software comercial.

Los nuevos estadísticos aprovechan la información suministrada por otros índices como:  $\chi^2_{LR}$ ,  $\chi^2_{GF}$ ,  $S_{LR}$ ,  $S_{GF}$ ,  $LR_{biv}$  y  $GF_{biv}$  y la emplean como elementos de cálculo, esto permite establecer una coherencia teórica y metodológica que se sustenta en la polémica sobre el tema, estableciendo una conexión con las reglas de interpretación de estos estadísticos, propuestas por diversos autores (Drasgow et al., 1995; Chernyshenko et al., 2001; Bartholomew et al., 2002; Bartholomew & Knott, 1999; Jöreskog & Moustaki, 2001), lo que supone una ampliación y mejora en la pirámide de ajuste de LVM.



Aunque estas medidas han sido planteadas para el tratamiento de datos ordinales desde la IRT utilizando FIML como método de estimación, los coeficientes  $\varphi_{LR}$  y  $\varphi_{GF}$  son independientes del método de estimación y pueden adoptarse con otros métodos de estimación como el MML. Esto supone ampliar su área de aplicación y la consideración de los mismos como medidas de ajuste de LVM en general.

Esta tesis plantea un conjunto de líneas de investigación futuras a seguir caracterizadas principalmente por:

- El desarrollo de procedimientos de software que permitan obtener índices específicos para evaluar LVM desde la IRT, que consideren el fenómeno del sparseness, tanto en modelos unidimensionales como multidimensionales, que contengan implementadas rutinas de simulación Monte Carlo para que permitan la validación posterior de estos índices
- La verificación del comportamiento de los coeficientes de ajuste  $\varphi_{LR}$  y  $\varphi_{GF}$  con ítems binarios.
- La comparación entre la implementación de Bootstrap paramétrico y la simulación Monte Carlo, en procedimientos de validación de los índices de ajuste.
- La utilización de los coeficientes de ajuste  $\varphi_{LR}$  y  $\varphi_{GF}$  para el ajuste de modelos politómicos de la TRI como el de respuesta graduada de Samejima y el de crédito parcial
- La comparación de los coeficientes de ajuste  $\varphi_{LR}$  y  $\varphi_{GF}$ , empleando FIML, MML u otros métodos de estimación.

En este sentido, se resalta que el tratamiento de datos ordinales en LVM está en pleno desarrollo y expansión y que cada día nos encontramos con nuevas estrategias que permiten aproximarnos con mayor precisión y rigor a las

realidades que pretendemos medir con nuestros datos y que por lo tanto el conocimiento específico de los mismos por parte de los investigadores puede resultar clave en la consolidación de los conocimientos derivados de la investigación.

## **Anexos**

# Tablas

1.1 Clasificación de los MVL según la métrica de variables latentes y observadas .....	12
2.1 Cargas estandarizadas y errores para la solución de un factor.....	44
2.2 Matriz de sumas de valores $\chi^2$ por pares de variables. Solución NOR de un factor.....	45
2.3 Matriz de sumas de valores $\chi^2$ por pares de variables. Solución POM de un factor. ....	46
2.4 Aproximación UV: índices de ajuste para la solución de un factor .....	47
2.5 Cargas factoriales estandarizadas y errores para la solución de dos factores.....	47
2.6 Matriz de sumas de valores $\chi^2$ por pares de variables. Solución NOR de dos factores.....	49
2.7 Matriz de sumas de valores $\chi^2$ por pares de variables. Solución POM de dos factores.....	50
2.8 Aproximación UV: índices de ajuste para la solución de dos factores.....	49
2.9 Aproximación UV: estadísticos descriptivos para cada índice de ajuste obtenidos en la simulación Monte Carlo.....	53
2.10 Aproximación IRT: estadísticos descriptivos para cada índice de ajuste obtenidos en la simulación Monte Carlo.....	53
3.1 Parámetros estimados e índices de ajustes modelo original de dos factores por método de estimación .....	66
3.2 Datos simulados: estadísticos descriptivos de índices de ajuste por método de estimación. ....	69
3.3 Resultados del ANOVA para los índices $LR_{uni}$ , $LR_{biv}$ , $GF_{uni}$ , $GF_{biv}$ .....	69
4.1 Procedimiento Monte Carlo: aproximaciones UV e IRT .....	80
4.2 Implementación Monte Carlo .....	81
4.3 Aproximaciones UV e IRT: medias y desviaciones típicas por procedimiento de simulación.....	84

5.1 Soluciones convergentes método de estimación con Bootstrap .....	97
5.2 Valores del coeficiente $\varphi_{LR}$ y $\varphi_{GF}$ con $\chi^2 m_i m_j = 3*5*5$ según la regla de interpretación .....	106

# Fíguas

2.1 Valores de los índices de ajuste por método de estimación .....	54
3.1 Diagramas de dispersión de cada índice por procedimiento de simulación.....	70
3.2 Diagramas de dispersión de índices por procedimiento de estimación .....	72
3.3 Valores de índices filtrados por Monte Carlo y POM.....	73
4.1 Aproximación UV: resultados de la simulación por método de estimación.....	85
4.2 Aproximación IRT: resultados de la simulación por método de Estimación .....	86
5.1 Pirámide de medidas de evaluación del ajuste en tablas de contingencia de amplia dispersión para modelos IRT .....	92
5.2 Evaluación del ajuste e índices de ajuste.....	98
5.3 Bootstrap: Distribución de los coeficientes $\varphi_{LR}$ y $\varphi_{GF}$ .....	102
5.4 Monte Carlo: Distribución de los coeficientes $\varphi_{LR}$ y $\varphi_{GF}$ .....	103
5.5 Valores del coeficiente $\varphi_{LR}$ y $\varphi_{GF}$ con $\chi^2_{mimj}=3*5*5$ para cada procedimiento de simulación.....	104
5.6 Valores del coeficiente $\varphi_{LR}$ y $\varphi_{GF}$ con $\chi^2_{mimj}=3*5*5$ para cada método de estimación .....	105

# Acrónimos

LVM	Latent Variable Models
UV	Underlying variables
IRT	Item response theory
LTM	Latent trait models
LCM	Latent class models
FA	Factor Analysis
SEM	Structural equation models
GLS	Generalized Least Squares
WLS	Weighted Least Squares
NOR	Normal function
POM	Proportional Odds Models
FIML	Full information maximum likelihood
MML	Marginal maximum likelihood
PM	Polychorics matrix
ACM	Asymptotic covariances matrix
OFA	Ordinal factor analysis
RMSEA	Root mean square error approximated
GFI	Global index fit
AGFI	Adjusted global index fit
CFI	Comparative index fit

# Notación

$\chi^2$	Chi cuadrado de Pearson	$\chi^2 = 2 \sum_{i=1}^{2^p} (O_i - E_i)^2 / E_i$ basado en frecuencias observadas y esperadas.
$G^2$	Chi-cuadrado Global Fits	$G^2 = 2 \sum_{i=1}^{2^p} O_i \log \left( \frac{O_i}{E_i} \right)$ basado en frecuencias observadas y esperadas.
$\chi_{LR}^2$	Chi-cuadrado likelihood ratio	$\chi_{LR}^2 = 2 \sum_r n_r \ln(p_r / \hat{\pi}_r)$ basado en la probabilidad de los patrones de respuesta
$\chi_{GF}^2$	Chi-cuadrado global fits	$\chi_{GF}^2 = \sum \left[ (n_r - N\pi_r)^2 / (N\pi_r) \right]$ basado en la probabilidad de los patrones de respuesta
$S_{LR}$	Suma de los $\chi_{LR}^2$ en tablas de contingencia	$S_{LR} = 2N \sum_{a=1}^{m_g} \sum_{b=1}^{mh} p_{ab}^{(gh)} \ln \left[ p_{ab}^{(gh)} / \hat{\pi}_{ab}^{(gh)} \right]$ para pares de variables.
$S_{GF}$	Suma de los $\chi_{GF}^2$ en tablas de contingencia	$S_{GF} = \sum_{a=1}^{m_g} \sum_{b=1}^{mh} (p_{ab}^{(gh)} - \hat{\pi}_{ab}^{(gh)})^2 / (\hat{\pi}_{ab}^{(gh)})$ para pares de variables.
$LR_{biv}$	Sumas de los $S_{LR}$ para $LR_{biv}$	$LR_{biv} = 2N \sum_{i=1}^{\frac{1}{2}n(n-1)} \left[ \sum_{a=1}^{m_g} \sum_{b=1}^{mh} p_{ab}^{(gh)} \ln \left( p_{ab}^{(gh)} / \hat{\pi}_{ab}^{(gh)} \right) \right]$ Por pares de variables de todo el modelo
$GF_{biv}$	Sumas de los $S_{GF}$ para $GF_{biv}$	$GF_{biv} = \sum_{l=1}^{\frac{1}{2}n(n-1)} \left[ \sum_{a=1}^{m_g} \sum_{b=1}^{mh} (p_{ab}^{(gh)} - \hat{\pi}_{ab}^{(gh)})^2 / (\hat{\pi}_{ab}^{(gh)}) \right]$ Por pares de variables de todo el modelo
$\varphi_{LR}$	Coeficiente de ajuste propuesto para $LR_{biv}$	$\varphi_{LR} = \frac{2LR_{fit}}{n(n-1)\chi^2 m_i m_j}$
$\varphi_{GF}$	Coeficiente de ajuste propuesto para $GF_{biv}$	$\varphi_{GF} = \frac{2GF_{fit}}{n(n-1)\chi^2 m_i m_j}$



## Referencias bibliográficas

Abad, F. J., Ponsoda, V., & Revuelta, J. (2006). Modelos politómicos de respuesta al ítem. Madrid: editorial La Muralla, S.A.

Ackerman, T. A. (2005). *Multidimensional Item Response Theory Modeling*. En: Maydeu-Olivares, A & McArdle, J. J., (eds), *Contemporary Psychometrics*. Lawrence Erlbaum Associates, Inc: New Yersey.

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley

Almond, G., & Verba, S. (1963). *The civic culture*. Princenton University Press. (*Translated to Spanish: La cultura cívica. Estudios sobre la participación política democrática en cinco naciones*. Madrid: Fundación Foessa (1970).

Alvarado , J. M., & Santisteban , C. (2001). Modelos psicométricos. Uned ediciones: Madrid.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness of fit indices for Maximun Likelihood Confirmatory Analysis. *Psychometrika*, 49, 155-173.

Anduiza, E., & Bosch, A. (2004). *Comportamiento político y electoral*. Editorial Ariel. Barcelona: España

Arbuckle, J. (1999). *Amos 4.0 User's Guide*. Chicago: Small Waters Corporation

Arminger, G., & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In R. Langeheine & J. Rost (Eds.) *Latent trait and latent class models* (51-73). New York: Plenum Press.

Bartholomew, D. J., & Leung, S.O. (2002). A goodness of fit test for sparse  $2^p$  contingency tables. *British Journal of Mathematical and Statistical Psychology*, 55, 1–15.

Bartholomew, D. J. (2002). Old and New approaches to Latent Variables Modelling. En: Marcoulides, G.A. & Moustaki, I. (eds). *Latent variable and latent structure models*. Lawrence Earlbaum Associates, Publishers: London.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London: Griffin.

Bartholomew, D. J., & Tzamourani, P. (1999). The goodness-of-fit of latent trait models in attitude measurement. *Sociological Methods and Research*, 27, 525–546.

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2002). *The Analysis and interpretation of Multivariate Data For Social Scientists*. Boca Ratón: Chapman and Hall-CRC.

Beaducel, A., & Wittmann, W. W (2005). Simulation study of fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling* 12(1), 45-75

Bearden, W. O., Sharma, S., & Teel, J. E (1982). Sample size effects on Chi Square and other statistics used in evaluating causal models. *Journal of Marketing Research*, 19 (November), 425-430.

Bock, R. D., Gibbson., R., & Muraki, E. (1988). Full-Information Item Factor Analysis. *Applied Psychological Measurement*, 12, 261-280.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46, 443-459

Bollen, K.A (1989). *Structural Equations with Latent Variables*. New York: Wiley

Bollen, K. A., & Stine, R. (1990). Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability. *Sociological Methodology*, Vol. 20. (1990), pp. 115-140.

Bollen, K., & Long, J. (1993). *Testing Structural Equation Models*. Londres: Sage Publications, Inc.

Bollen, K.A., & Stine, R.A. (1993). *Bootstrapping goodness-of-fit measures in structural equation models*. In: Bollen, K.A. y Long, J.S. (eds). *Testing Structural Equation Models*. Sage Publications, Beverly Hills, CA.

Bolt, D.M. (2005). Limited and full information estimation of item response theory models. En: Maydeu-O, A. & McArdle, J. J.(eds). *Contemporary*

psychometrics. Latent variable and latent structure model. Lawrence Erlbaum Associates, Publishers: London.

Bone, P. F., Sharma, S., & Shimp, T. A (1989). A Bootstrap procedure for evaluating goodness-of-fit. *Journal of Marketing Research*, 26,(1),105-111.

Boomsma, A. (1982). *The robustness of LISREL Against small sample size in Factor Analysis Models*. In Systems under indirect observation: causality, structure, prediction, Jöreskog, K. G. & Wold, H (eds). Amsterdam: North-Holland Publishing Company

Brady, H. (1999). *Political Participation*. En: Robinson, J.P., Shaver, P.R. & Wrightsman, L.S. Measures of Political Attitudes. San Diego: Academic Press

Browne, M. W. (1982). Covariance structure. En: Hawkins, D.M. (eds). En Topics in Applied Multivariate Analysis. Cambridge University Press: Cambridge.

Browne, M. W. (1984). Asymptotic distribution free methods in análisis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83

Cai, L., Maydeu-Olivares, A., Coffman, D., & Thissen, D (2006). Limited-information goodness-of-fit testing of item response theory models for sparse 2<sup>p</sup> tables. *British Journal of Mathematical and Statistical Psychology* , 55, 1–15.

Carton, A., Van Geel, H., & De Pelsemaeker, S. (2004). *Basis documentatie: Social-culturele verschuivingen in Vlaanderen*. Brussel: MVG – APS.

Centro de Investigaciones Sociológicas de España, (2004). *Ciudadanía y Participación (ISSP)*. Estudio N° 2575. Madrid: CIS.

Chan, W & Chan, D. W. (2004). Bootstrap Standard Error and Confidence Intervals for the Correlation Corrected for Range Restriction: A Simulation Study. *Psychological Methods*, 9 (3), 369-385.

Chernick, M. R. (1999). Bootstrap Methods: A Paractioner`s Guide. New Cork:Jhon Wiley & Sons, Inc.

Chernyshenko, O. S., Stark, S., Chan, K., Drasgow, F., & Williams, B. (2001). Fitting Item Response Theory Models to Two Personality Inventories: Issues and Insights. *Multivariate Behavioral Research*, 36 (4), 523-562

Collins, L M., Fidler, P.L., Wugalter, S. E., & Long, J. D. (1993). Goodness-of-fit Testing for latent Class models, *Multivariate Behavioral Research* 29(3): 375-389.

Craig, K. E (1992). Applying the Bollen-Stine Bootstrap for Goodness-of-Fit Measures to Structural Equation Models with Missing Data. *Multivariate Behavioral Research*, 37(3), 359-377

D'Agostino, R. B. (1986). Test for the normal distribution. En: D'Agostino, R. B. & Stephens, M. A. (Eds). *Goodness of fit techniques* (367-419). New York: Marcel Dekker.

Davey, A., Savla, J., & Luo, Z. (2005). Issues in Evaluating Model Fit With Missing Data. *Structural Equation Modeling*, 12(4), 578-597.

Drasgow, F., Levine M. V., Tsien, S., Williams B. A., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19, 143-165.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1-2

Efron, B. (1982). *The Jackknife, the Bootstrap, and other Resampling Plans*. Philadelphia: SIAM.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171-200.

Efron, B. (1988). Bootstrap confidence intervals: good or bad?. *Psychological Bulletin*, 104, 293-296.

Efron, B. (1990). More Efficient Bootstrap Computations. *Journal of American Statistics Association*, 85, pp. 79-89.

Efron, B. & Tibshirani, R.J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1), 54-77.

Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Nueva York: Chapman & Hall.

Everitt, B. S (1984). *An introduction to Latent Variable Models*. Chapman and Hall: London .

Fan, X. (1994). *Does Bootstrap Procedure Biased Estimates? An Empirical Examination for a Case of Multiple Regression*. Paper presented at the Annual Meeting of AERA, New Orleans, Abril.

Fleishman, A. I. (1978) A method for simulating non-normal distributions. *Psychometrika*, 43, 521-532.

Gerbing, D. W., & Anderson, J. C.(1993). Monte Carlo evaluations of goodness of fit indices for structural equations models. En: Bollen, K, A. & Long, J.S. (eds). *Testing structural equation models*. Sage publicactions, Inc: California.

Hartmann, W. H (2005). *Resampling Methods in Structural Equation Modeling*. In: Maydeu-Olivares, A & McArdle, J. J.,(eds), *Contemporary Psychometrics*. Lawrence Erlbaum Associates, Inc: New Yersey.

Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oakes: Sage Publications.

Jhonson, V. E., & Albert, J. H. (1999). *Ordinal data modeling*. Springer: New York.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183-202.

Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, 24, 387-404.

Jöreskog, K.G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, 59, 381–389.

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36, 347-387.

Jöreskog, K. G., & Moustaki, I. (2006). *Factor analysis of ordinal variables with Full Information Maximun Likelihood*. Disponible en: <http://www.ssicentral.com/lisrel/techdocs/ordfiml.pdf>

Jöreskog, K. G., & Sörbom, D. (2002). *PRELIS 2 User`s refrence guide*. Lincolnwood: Scientific Software International, Inc.

Jöreskog, K. G. & Sörbom, D. (2006 a). *LISREL 8.81*. Lincolnwood: Scientific Software International, Inc.

Kelderman, H. (1996). Multidimensional Rasch models for partial-credit scoring. *Applied Psychological Measurement*, 20, 155-168

Langeheine, R., Pannekoek, J., & Van de Pol, F. (1996). Bootstrapping Goodness-of-fit measures in Categorical Data Analysis. *Sociological Methods and Research* 24 (4): 492-516

Lee, S. Y., Poon, W. Y., & Bentler, P. (1990). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters*, 9, 91-97.

Lee, S. Y., Poon, W. Y., & Bentler, P. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, 57, 89-105.

Loehlin, J. C. (2004). *Latent variable models*. Lawrence Erlbaum Associates, Inc: New Jersey.

Marsh, H., & Balla, J. (1994). Goodness of fit in confirmatory factor analysis: the effects of sample size and model parsimony. *Quality & Quantity*, 28, pp.185-187

Maydeu-Olivares, A. (2005). *Linear Item Response Theory, Nonlinear Item Response Theory and factor Analysis: A unified framework*. En: Maydeu-Olivares, A & McArdle, J. J. (Eds). *Contemporary Psychometrics*. Lawrence Erlbaum Associates, Inc: New Jersey.

Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*, 71, 57-77



Mavridis, D., Moustaki, I & Knott, M. (2007). *Goodness of fit Measures for Latent Variable Models for Binary Data*. En: Lee, S. K (Ed). Handbook of latent variable and related models. North- Nolland: Elsevier

Meade, A. W & Lautenschlager, G. J. (2004). A Monte Carlo Study of Confirmatory Factor Analytic Test of Measurement Equivalence/Invariance. *Structural Equation Modeling*, 11, 60-72

Money, C. Z. (1997). *Monte Carlo simulation*. Thousands Oaks, CA: Sage

Moustaki, I. (2000). A Latent Variable Model for Ordinal Variables. *Applied Psychological Measurement* , 24, 211-223

Moustaki, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology* , 56, 337–357

Moustaki, I., & Knott, M. (2005). *Computational aspects of the E-M and Bayesian Estimation in Latent Variable Models*. En: Van der Ark, L.A. & Marcel, A.C. & Sijtsma, K(eds). New Developments in Categorical Data Analysis for the Social and Behavioral Sciences. Lawrence Earlbaum Associates, Publishers: London.

Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, 65, 391-411.

Moustaki, I., Joreskog, K.G., & Mavridis, D. (2004). Factor Models for Ordinal Variables With Covariate Effects on the Manifest and Latent Variables: A Comparison of LISREL and IRT Approaches. *Structural Equation Modeling*, 11(4). 487-513.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, 14, 59–71.

Muraki, E., & Carlson, E. (1995). Full-information factor analysis for polytomous item Responses. *Applied Psychological Measurement*, 19, 73–90.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variables indicators. *Psychometrika*, 49, 115–132.

Ostini, R & Nering M. L. (2006). *Polytomous item response theory models*. Thousands Oaks, CA: Sage.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J. & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8, 287–312.

Psychometrika (1979). Publication policy regarding Monte Carlo studies. *Psychometrika*, 44, 133–134.

Read, T. R. C & Cressie, N. A. C. (1988). Goodness-of-Fit statistics for discrete Multivariate data. New York: Springer – Verlag

Reiser, M., & Lin, Y. (1999). A Goodness-of-Fit Test for the Latent Class Model When Expected Frequencies are Small. *Sociological Methodology*, 29, pp. 81–111.

Revuelta, J., & Ponsoda, V. (2003). *Simulación de modelos estadísticos en las ciencias sociales*. Madrid: La Muralla.

Robert, C .P., & Casella, G. (2004) *Monte Carlo methods*. New York: Springer.

Rubenstein, R. Y. (1981). *Simulation and the Monte Carlo method*. New York: John Wiley.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 17.

Samejima, F. (1997). Graded response model. In Van der Linden, W.J., & Hambleton, R.K. (Eds). *Handbook of modern item response theory* (85-100). New York: Springer.

Santisteban , C (1990). Psicometría. Teoría y práctica en la construcción de test. Ediciones Norma: Madrid.

Schrage, L. (1979). A more portable FORTRAN random number generator. *ACM Transactions on Mathematical Software*, 5, 132-138.

Sclove, S. (1987). Application of model-selection-criteria to some problems of multivariate analysis. *Psychometrika*, 52, 333-343.

Skron dal, A. & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling*. Boca Ratón: Chapman and Hall-CRC.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology* 15, 201-293.

Stapleton, L. M. (2006). An Assessment of Practical Solutions for Structural Equation Modeling with Complex Sample Data. *Structural Equation Modeling* 13(1), 28-58

Stephenson, M. T. & Holbert, R.L. (2003). A Monte Carlo Simulation of Observable Versus Latent Variable Structural Equation Modeling Techniques. *Communication Research*, 30(3), 332-354.

Tanaka, J. S. (1987).How Big is Big Enough?: Sample Size and Goodness of Fit in Structural Equation Models with Latent Variables. *Child Development*, 58, pp. 134-146.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one parameter logistic model. *Psychometrika*, 47, 175-186.

Vale, C. D., & Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48, 465-471

Wiggins, R. D., & Sacker, A. (2002). Strategies for handling missing data in SEM: a user's perspective. En: Marcoulides, G. & Moustaki, I (eds). *Latent variable and latent structure model*. Lawrence Erlbaum Associates, Publishers: London.

Wilson S.; Meagher G.; Gibson R.; Denemark, D & Western, M. (2005). *Australian Social Attitudes: The First Report*. UNSW Press. Australia

Ximénez, C., & García, A (2005).Comparación de los métodos de estimación de máxima verosimilitud y mínimos cuadrados no ponderados en el análisis factorial confirmatorio mediante simulación Monte Carlo. *Psichotema*, 17 (3), 528-535

Young, G. A (1994). Bootstrap: More than a Stab in the Dark?. *Statistical Science*, 9,3, pp. 382-395.